Ubiquitin binding by the p62 UBA domain

Jennifer Adlington, BSc (Hons)

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy,

July 2013

Abstract

Over 30 different mutations in p62 UBA have been identified in patients with Paget's disease of bone (PDB). The mechanisms which underlie PDB are poorly understood, although impaired ubiquitin binding has been identified as a mechanism in the onset of disease. However, the decrease in affinity is subtle for many PDB mutants. The p62 UBA is unique amongst UBAs since it exists as a highly stable dimer but binds to ubiquitin as a monomer. The dimerisation interface partially occludes the ubiquitin binding interface resulting in competing equilibria. The factors which regulate the affinity of p62 UBA were examined in this thesis.

In isolated p62 UBA the monomer-dimer equilibrium and the effects of phosphorylation were investigated. By mutating residues at the dimerisation interface, weaker dimers which had a higher affinity for ubiquitin were produced. The weak dimers had an increased population of monomer at equilibrium. A phosphorylation site at Ser403 in p62 UBA was recently identified. Phosphomimetic mutants which showed subtle increases in affinity for ubiquitin were generated. The increase was attributed to the close proximity of Ser403 to the ubiquitin binding interface.

Factors outside the UBA also have a role in regulating UBA affinity. Binding by p62 UBA was therefore probed in longer p62 constructs. A fragment of p62 encoding residues 300-440 was used to investigate p62 binding to multiple proteins. A ternary complex was formed, but an allosteric relationship was not observed by ubiquitin and MAP-LC3 in binding to the p62 fragment. A model of full length oligomeric p62 was generated to probe avid binding to polyubiquitin chains. The model showed higher affinity for linear diubiquitin than monoubiquitin suggesting avidity effects are influential in oligomeric p62 binding. Since the effects of PDB mutations and phosphorylation are subtle, they are likely to be amplified by avidity *in vivo*.

Acknowledgements

First and foremost I would like to thank my supervisor Prof Mark Searle for giving me with the opportunity to work on this project. I would also like to thank him for his constant support and inspirational ideas. I would also like to thank our collaborator Dr Robert Layfield who has provided fantastic insights to the biological side of the project. I also thank the University of Nottingham Department of Chemistry and the BBSRC for funding.

I would also like to give thanks to all members of the Searle and Layfield groups who had previously worked on the project to provide a great starting point for my PhD. Much of the structural and biophysical data on the wild type p62 UBA was completed by Dr Jed Long and Dr Thomas Garner, who I extend a special thanks to. Despite crossing the Atlantic, Tom has ensured he is never an email away and for that I am eternally grateful.

I would also like to acknowledge the members of the Searle group past and present for taking the time to train me in the lab techniques and for answering all of my questions. In particular the students who began their PhD journeys with me Liz Morris and John Edwards, I could not have done this without either of you. I also thank Dr Huw Williams who has been especially patient when teaching a molecular biologist how to understand protein NMR.

A particular mention must go to the friends who have made my time in Nottingham so memorable: Emily Golding, Jill Sollenberger, Chris Collier, Olivier Rannou, Matt Green, Dr Geoff Briggs, Jose Afonso, Matt Jenner, Kleitos Sokratous, Dr Huey-Jen Fang, Dr Shalu Sharma-Kharkwal, Lei Zhang, Alice Goode, JP Phillips and Dr Pallavi Sharma. I also thank Dr John Moses for always being there to support me when the times were tough.

I am forever grateful to my best friend and long suffering housemate Kerry Monk, who has put up with me for nearly the entirety of my PhD. Your friendship and ability to keep things in perspective has been invaluable. I also thank my wonderful Mum for believing in me. Finally a special thank you goes to Michael Jones who supported my decision to move away. I am so grateful for your love and endless support, I dedicate this thesis to you.

Table of contents

Abbreviations

A_{260}	Absorbance at 260 nm
A_{280}	Absorbance at 280 nm
aPKC	atypical protein kinase C
APS	Ammonium persulfate
BMRB	Biological magnetic resonance data bank
bp	base pairs
BSA	Bovine serum albumin
CD	Circular Dichroism
CSP	Chemical shift perturbations
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
DUB	Deubiquitinase
E. Coli	*Escherichia coli*
EDTA	Ethylene diamine tetraacetic acid
ESI-MS	Electrospray Ionisation Mass Spectrometry
GST	Glutathione S-transferase
HSQC	Heteronuclear single quantum coherence
IκB	Inhibitor of κB
IKK	IκB kinase
IPTG	Isopropyl-β-D-Thiogalactopyranoside
ITC	Isothermal titration calorimetry
kDa	Kilodaltons
LB	LuriaBroth
M	Molar
m/z	Mass/charge ratio
NEMO	NF-κB essential modulator
NF- κB	Nuclear Factor κB
NMR	Nuclear Magnetic Resonance

NOE	Nuclear overhauser effect
NOESY	Nucelar overhauser effect spectroscopy
OD_{595}	Optical density at 595 nm
PAGE	Polyacrylamide gel electrophoresis
PB1	Phox and Bem1p
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PDB	Paget's disease of bone
ppm	Parts per million
PTM	Post translational modification
RANK	Receptor activator of NF-κB
RANKL	Receptor activator of NF-κB ligand
RDC	Residual dipolar couplings
rpm	revolutions per minute
S.Cerevisiae	*Saccharomyces Cerevisiae*
SDS	Sodium dodecyl sulphate
TEMED	*N,N,N,N*-Tetramethylethylethylendiamine
TNFα	Tumor necrosis factor alpha
T_m	Midpoint of unfolding
TOCSY	Total correlation spectroscopy
TRAF-6	TNF receptor associated factor 6
TROSY	Transverse relaxation optimised spectroscopy
Tris.HCl	tris(hydroxymethyl)aminomethane HCl
UBA	Ubiquitin-associated domain
UBAN	ubiquitin binding in ABIN and NEMO
UBD	Ubiquitin binding domain
UPS	Ubiquitin Proteasome System
UV	Ultraviolet
wt	Wild type

Amino acid abbreviations

Amino acid	3 letter code	1 letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histadine	His	H
Isoleucine	Iso	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1.0 Introduction

1.1 Cell signaling pathways

There are an estimated 1 billion individual protein molecules in a typical mammalian cell[1]. Each protein performs a specific function within the cell. With such a diverse range of functions performed by a vast network of molecules, a communication mechanism enabling each protein to reach its correct binding partner is required. Thus, complex signalling pathways which utilise a relatively low number of proteins whilst maintaining a high degree of specificity have evolved. Over the past twenty years theories have emerged about proteins which belong to certain signalling pathways achieving their specificity by being sequestered to certain compartments to form a spacially organised cell[2,3]. This organisation can be achieved using subcellular organelles, co localisation to the cell membrane or scaffold proteins[4]. Cell signalling pathways are activated by various stimuli often resulting in changes in gene expression or gene splicing. This enables the signalling pathway to regulate the levels of the different proteins involved in that pathway to generate the desired response. In addition to cellular organisation, the proteins themselves, are spacially organised keeping their catalytic domains separate from other domains which might be involved in targeting them to specific cellular compartments. Overall the evolution of these pathways to convey molecular signals with such high specificity is remarkable.

1.2 Post translational modifications

Post translational modifications (PTMs) are chemical modifications of proteins which occur after protein translation. Nearly every protein that is synthesized by the cell undergoes a form of PTM in its lifespan. Examples of chemical modifications include acetylation and methylation, which alter the charge of amino acids, and lipidation, which enable proteins to be anchored to cell membranes. More complex PTMs include phosphorylation and ubiquitination, both of which are reversible covalent attachments. In phosphorylation proteins are modified by phosphate groups, whereas ubiquitination involves the addition of the small protein ubiquitin to act as a molecular signal. Protein phosphorylation is a vital regulatory mechanism as many proteins become activated or deactivated by the modification. Both phosphorylation and ubiquitination processes have evolved families of enzymes for the attachment or the removal of the chemical modification. In the case of phosphorylation protein kinases attach phosphate groups to amino acids and phosphatases remove phosphate groups. Ubiquitination involves a cascade of enzymes for its attachment and deubiquitinases (DUBs) for its removal, both of which are discussed in detail later in this thesis. However there are two vast differences between ubiquitination and phosphorylation: the size of the attachment and the ability of ubiquitin to form long polymers[5]. It is the added complexity of protein ubiquitination which has left the ubiquitin system with the greatest challenges to fully understand.

1.3 Ubiquitination

Ubiquitin was discovered in 1975 by Aaron Cienhanover, Avram Hershko and Irwin Rose. Their discovery was revealed to be so monumental that the trio were awarded the Nobel Prize in Chemistry in 2004[6]. Ubiquitin is a small protein with a molecular weight of just 8.5 kilo daltons (kDa) and a surface area of 4800 $Å^2$. It consists of 76 amino acids that adopt a β-grasp globular fold composed of a 5 stranded β-sheet flanked by a single α-helix[7,8]. The backbone of ubiquitin is rigid but conformational flexibility has been shown in the β1-β2 loop (residues 8-11) and the C-terminal (residues 72-76)[9]. This protein is expressed ubiquitously in eukaryotes, hence it's name. Ubiquitin has a highly conserved protein fold maintained by the exceptionally high percentage sequence identity found amongst various orthologues[10]. This is demonstrated well by yeast and human orthologues, which differ by just 3 residues, suggesting a high evoluntionary pressure to conserve the structure of ubiquitin. The ubiquitin fold has also been adopted by domains referred to as ubiquitin-like (UbL) domains. Prokaryotes lack a functionally equivalent protein; however, proteins containing the ubiquitin fold exist highlighting the evolutionary precursors to ubiquitin in these cells.

Ubiquitin is covalently attached to proteins as a PTM to act as a molecular signal to determine the fate of the target protein. In protein ubiquitination Lys residues in the target protein become linked to the C-terminal diglycine motif of ubiquitin. The resultant isopeptide bond is formed between the carboxy-terminus of ubiquitin and the ε-amino group of a Lys residue in a target protein. Ubiquitin itself can also be ubiquitinated permitting the formation of long polyubiquitin chains via any one of its seven Lys residues which serve as additional points of attachment[11]. Linear polyubiquitin chains can also be formed with a peptide bond being formed by the carboxy-terminus and amino-terminus of two ubiquitin molecules. Ubiquitin was initially discovered for its ability to mark proteins for degradation by the proteasome in a system termed the ubiquitin-proteasome system (UPS)[12]. However extensive study since the discovery of the UPS has revealed a complicated system which exploits polymerisation to modulate various molecular signals[5]. The complexity of this system has led to thoughts on the level of protein ubiquitination and chain linkage becoming the key to cracking the ubiquitin code.

Unlike most proteins ubiquitin has a remarkable ability to bind with high specificity and sufficient affinity to a diverse array of binding partners. Most residues in ubiquitin are rigid; however, the C-terminal tail exhibits some conformational flexibility[8]. Recently residues in the β1/β2 loop, which contains Leu8, were shown to be flexible[13]. The study used relaxation dispersion nuclear magnetic resonance (NMR), refined by residual dipolar couplings (RDC's), to investigate the full dynamic behaviour of ubiquitin in solution. The study revealed a pincer-like motion of ubiquitin contributing to the molecular recognition of binding partners[13].

Ubiquitin has several binding interfaces, the most common of them being centred on Leu8, Ile44 and Val70[14]. However non canonical binding sites have also been identified on the surface of ubiquitin[15]. The presence of multiple binding sites permits the formation of ubiquitin-mediated ternary complexes, in which ubiquitin can be transferred between proteins using a 'hand off' mechanism[16]. Other binding surfaces centred on His68, Glu58 and C-terminal tail have been identified, although the binding site of His58 is often combined with the canonical Ile44 patch as they are located in such close proximity. Nearly all of the binding sites identified on ubiquitin are located n the same face of the molecule (figure 1.1). Both conformational selection and induced fit mechanisms have been implicated in the binding of ubiquitin to other proteins[13, 17].

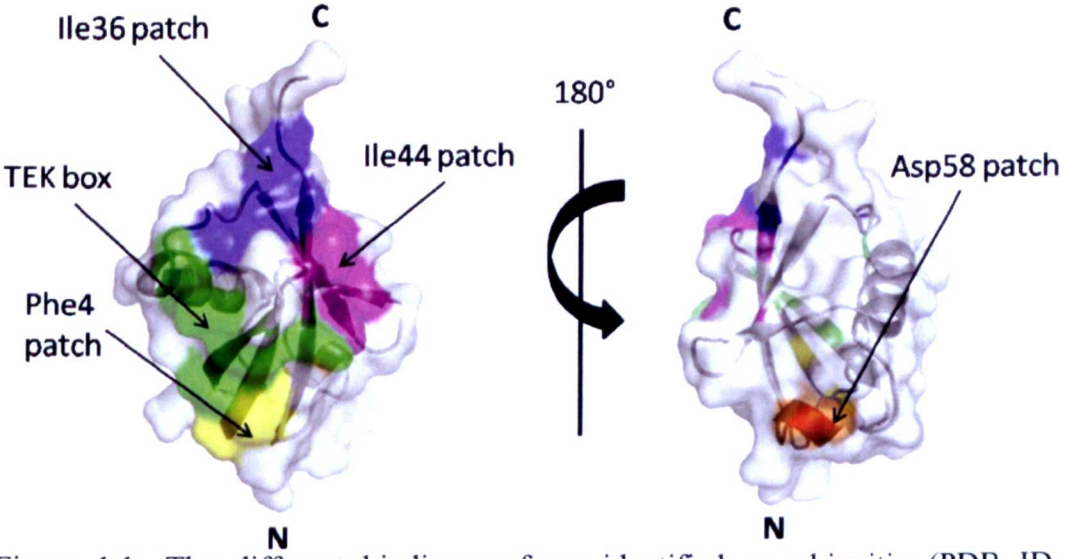


Figure 1.1- The different binding surfaces identified on ubiquitin (PDB ID 1UBQ). The known patches are coloured and labelled. Nearly all of the binding sites are located on one face of ubiquitin, with only the Asp58 patch located on the other face.

1.3.1 Ubiquitination enzymes

The covalent attachment of ubiquitin to target proteins is completed using different classes of enzymes (figure 1.2). First a ubiquitin-conjugating enzyme, known as an E1 enzyme, adenylates the C-terminal glycine to form an intermediate which is then transferred to a cysteine residue in the E1. The resultant ubiquitin-E1 thiol ester is termed activated ubiquitin and its formation is ATP dependent[11, 12]. There are only 2 E1 enzymes encoded in the human genome which have been identified to date, UBE1 and UBE1L2. It had been previously assumed that a single E1, UBE1, performed the initial step in the cascade alone, until the discovery of UBE1L2 in recent years[18,19]. The two enzymes share 40 % sequence identity. Other E1 enzymes exist for the attachment of UbL proteins but are not able to activate ubiquitin itself[20].

The next step involves a ubiquitin-conjugating enzyme, known as an E2, to catalyse the transfer of the activated ubiquitin from the cysteine of the E1 to the active site cysteine in the E2. The reaction is a transthioesterification reaction. There are significantly more E2 enzymes than E1 enzymes. The human genome encodes around thirty seven different E2's[21]. A study using E2 enzymes indicated that the E2's have the most influence on which of the seven Lys residues in ubiquitin is targeted by polyubiquitination. It highlighted how E2 enzymes can only produce polyubiquitin chains targeting either a subset of Lys residues or an explicit Lys residue. Further experiments which included the E3 ubiquitin ligase, the final enzyme involved in the ubiquitination process, showed that the E2's capability to polymerise through a specific Lys residue is not altered by the presence of the E3[22].

As briefly mentioned, the final step in the attachment of ubiquitin to the target molecule is for a ubiquitin ligase known as an E3, to facilitate isopeptide bond formation between the two proteins. The mechanism of action of E3s is poorly understood. However it is believed the E3 binds to the E2 ubiquitin thioester and activates release of ubiquitin[21]. There are an extraordinary number of E3s, in fact more than 600 have been identified to date[23]. This exceptional number reflects the large array of substrates targeted for ubiquitination. These enzymes can be divided into 3 different families of proteins, Really Interesting New Protein (RING), Homologous to E6AP Carboxy Terminus (HECT) and UFD2 homology (U-box)

proteins. RING E3s contain 2 co-ordinated zinc ions which are surrounded by histidine and cysteine residues to form a zinc finger motif[24].

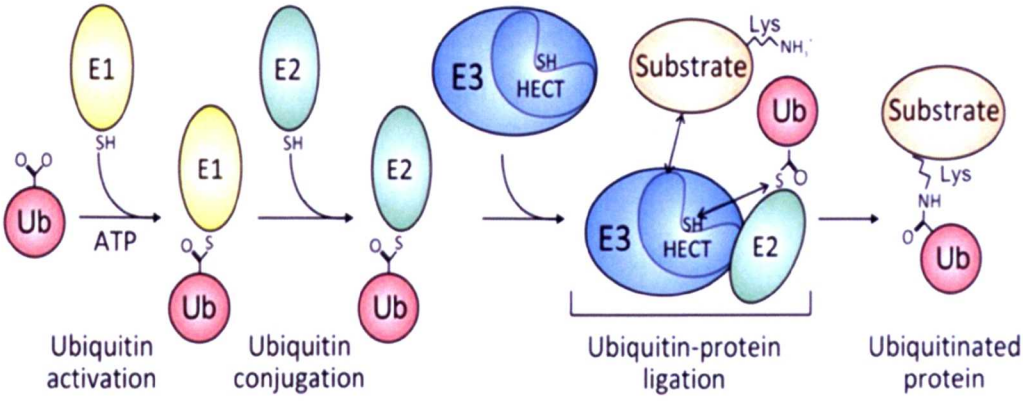


Figure 1.2– A schematic diagram of the ubiquitination cascade. The E1, E2 and E3 enzymes attach the ubiquitin moiety to the substrate via a Lys residue in the ubiquitin.

More recently a group of enzymes termed E4's were discovered in 1999[25]. Not all ubiquitinations are catalysed by an E4 but they do represent a separate protein family in their own right. E4 enzymes propagate chain extension for proteins already primed for degradation, leading to the formation of long polyubiquitin chains[26]. Much of the focus on E4 enzymes is linked to the first E4 to be identified, a protein known as UFD2[25, 27, 28].

1.3.2 Types of ubiquitination

The complexity associated with the ubiquitin system is caused by the ability of ubiquitin to polymerise[5]. As discussed, the polymerisation can occur through any on the seven Lys residues found in ubiquitin or via the N and C terminals of consecutive ubiquitin molecules. The ubiquitin modification can be a monoubiquitination, a multi-monoubiquitination or a polyubiquitination. Polyubiquitin chains can be either homotypic (i.e. linked via the same Lys residue in adjacent ubiquitin molecules or linear polyubiquitin) or heterotypic (i.e. linked via different Lys residues to form chains of mixed linkage). Heterotypic chains can also contain certain ubiquitin molecules which have become multiubiqutinated to form forked chains. The exploration of homotypic chains has been underway for the past few decades since the first structure of a diubiquitin

chain was reported[29]. However, heterotypic chains present a bigger challenge and research into these chains is rather limited.

Monoubiquitin is the simplest form of ubiquitination, but like polyubiquitination it has been implicated in a variety of cellular processes including DNA repair and viral budding[30,31]. Multi-monoubiquitination is also a common modification to proteins. The p53 protein which has been implicated in several cancers is regulated by the E3 mdm2 which has been shown to favour multi-monoubiquitination over polyubiquitination[32]. Although the simplest form of ubiquitination the enzymes which catalyse monubiquitination have been shown to exert a high degree of specificity. These enzymes have to recognise substrate Lys residues whilst sparing those on ubiquitin from modification.

The Lys residues on ubiquitin are all surface exposed. They are located all around the ubiquitin molecule and are not confined to a specific patch (figure 1.3). Consequently, when polymerised the ubiquitin chain can adopt diverse structures with differing conformational flexibilities. The topologies of the different chains cause different interaction surfaces to be presented to other proteins and different flexibilities of the linker between the two moieties. The hydrophobic patch centred on Ile44 can be fully or partially buried or exposed in different orientations to enable high affinity binding to certain linkages. These residual differences account for the specific molecular recognition of the chain as a distinct molecular signal. The surfaces are recognised by different ubiquitin binding domains (UBDs) and DUBs with certain linkages preferentially bound. Many structural investigations using both X-ray crystallography and solution NMR have led to multiple conformations of diubiquitin chains formed by the same linkage[33,34,35, 36]. This has lead to speculation about ubiquitin chains being viewed as flexible units which can be shaped by their binding partner.

The diversity of polyubiquitin chains has attracted the interest of synthetic chemists in recent years. The synthesis of all possible diubiquitin chains as synthetic analogues has now been completed by various groups[37,38,39]. The ability to probe the selectivity of various UBDs, including DUBs, can be achieved using such synthetic analogues. This approach could provide substantial advances in the

field as ubiquitination enzymes for the synthesis of Lys48, Lys63, Lys11 and linear chains are only currently known.

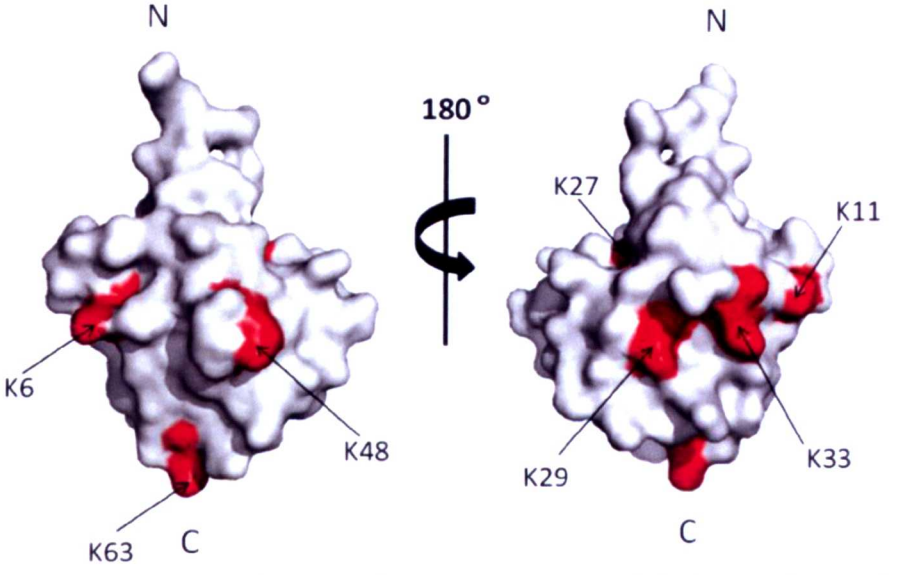


Figure 1.3 – Ubiquitin surface plot (PDB ID 1UBQ) which shows the locations of the seven Lys residues (highlighted in red). The Lys residues are used as points of attachment in the formation of polyubiquitin chains. An isopeptide bond is formed between the carboxy-terminal of ubiquitin and the ε-amino group of the Lys residues on another ubiquitin molecule. The N and C terminals (also indicated) in consecutive ubiquitin molecules can also form peptide bonds to form linear polyubiquitin chains.

1.3.3 Ubiquitin chain linkage

A recent study using *S. Cerevisiae* profiled the yeast proteome and found evidence for all seven Lys linked polyubiquitin chains and linear chains [40]. This observation had previously not been confirmed *in vivo* and also highlighted some intriguing results. The most abundant linkage was unsurprisingly found to be Lys48 at 29% with the unconventional Lys11 linakge highlighted as the second most frequent linkage at 28 %. The third most common linkage was found to be Lys63 at 17%. The roles of both Lys48 and Lys63 linked polyubiquitin chains have now been clearly established[41,42]; however, the revelation of the abundance of Lys11 chains *in vivo* has provoked extensive research into the role of these chains[33]. The structures of many of diubiquitins are now known, revealing differences in the overall structure, the flexibilities of the linkers and the orientations of the binding interfaces (figure 1.5).

1.3.3.1 Lys48 linkages

Lys48 chains are the most extensively studied chains because they were the first linkage to be discovered. The ubiquitination reaction for the formation of Lys48 linked chains *in vitro* has been published allowing chains of this linkage to be commercially available. Such chains are constructed by the E2 enzymes UBE2R1, UBE2G2 and UBE2K[43,44]. Lys48 linked tetraubiquitin is the minimum requirement for degradation by the 26S proteasome[41]. The canonical Leu8, Ile44, Val70 hydrophobic patch on ubiquitin is recognised by the proteasome[45,46]. The tag marks the target protein for degradation thus allowing the amino acid components to be recycled by the cell during protein synthesis. Ubiquitin itself is also recycled but it is not degraded by the proteasome. The proteasome regulates the cellular concentrations of proteins and therefore represents a vital degradation pathway. The crystal and solution NMR structures of Lys48 linked diubiquitin[29,47] and tetraubiquitin[45, 46, 48,49] have provided great insight to the specific conformation of this linkage. The 2-fold symmetry in each of the ubiquitin pairs in Lys48 tetraubiquitin is quite different from the 2-fold symmetry observed in the structure of Lys48 linked diubiquitin. There are multiple hydrophobic contacts in Lys48 tetraubiquitin causing the formation of a buried hydrophobic core (figure 1.4). This hydrophobic surface is located centrally to the diubiquitin structure and is more open than in Lys48 tetraubiquitin. The hydrophobic interface is therefore liberated in Lys48 linked diubiquitin, freeing it to interact with various chain-recognition factors, including the proteasome[50-52].

Solution NMR studies aimed at probing discrepancies between different crystal structures have highlighted that Lys48 chains are inheritantly flexible, with a pH dependent equilibrium between open and closed conformations existing[53]. The compact conformation is dominant at physiological pH (pH 7), with the subunits forming stabilising hydrophobic contacts with one another burying the canonical binding site. During an interaction, the closed conformation opens due to the weakness of the interaction between the subunits, presenting the hydrophobic patches. In Lys48 diubiquitin the proximal ubiquitin (the subunit which carries the free Gly76 residue) binds more strongly than the distal molecule enabling the UBD to position itself between the two subunits forming a sandwich-like complex[54]. In the open conformation, which is more populated at low pH (pH

4.5), there are no hydrophobic contacts between the two subunits exposing the patch for interaction with UBDs and the proteasome[53]. More recently, this conformation has been shown to interact with UBDs in the same manner and with comparable affinity as the closed conformations, highlighting a conserved binding mechanism amongst Lys48 linked chains[49].

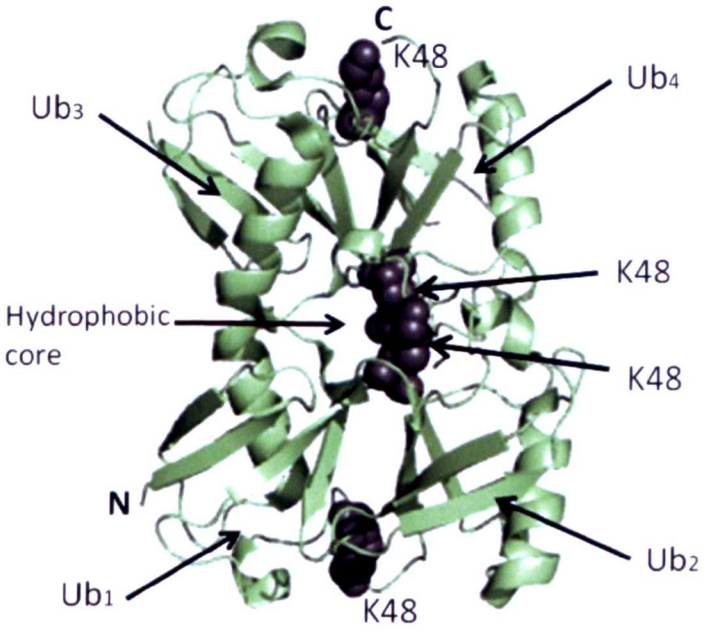


Figure 1.4 – The structure of Lys48 linked tetraubiquitin (PDB ID 2O6V).The Lys48 residues are highlighted as dark grey spheres. The structure folds back on itself forming contacts between the individual subunits to adopt a compact overall structure with a buried hydrophobic core.

1.3.3.2 Lys63 linkages

Lys63 linked polyubiquitin chains are assembled by UBE2N and UBE2V1 E2 and E3 enzmyes[55,56]. These chains, unlike Lys48 chains, form open and extended conformations in which there are no contacts formed between the ubiquitin subunits. The hydrophobic binding patch centred on Ile44 is surface exposed in this open conformation, even in chains with more than two ubiquitin moieties[57,58,35,59]. These chains can rotate without restraints but can also form more compact structures due to the flexibility of the linker region between the ubiquitin moieties. Not only is the structure of Lys63 chains completely different to Lys48 chains but so is the function. Lys63 linked polyubiquitin chains are not linked to proteasomal degradation[60,42]. Lys63 linked chains have been linked to several diverse functions ranging from DNA repair to cell signalling and

trafficking and most importantly to this study, nuclear factor-κB (NF-κB) activation[42,61]. Lys63 linked chains have been shown to exhibit structural flexibility depending on binding partner. When bound to the associated molecule with the Src homology 3 (AMSH) domain Lys63 linked diubiquitin is stretched[62]; whereas, in complex with an antibody a more compact structure with contacts between the subunits is utilised[63].

1.3.3.3 Lys11 linkages

A recent advancement in the ubiquitin field is the discovery of the high abundance of Lys11 linked chains in the yeast proteome[40]. Although similar studies in higher order eukaryotes did not yield similar results[64]. The anaphase-promoting complex (APC/C) has been identified as an E3 which assembles Lys11 linked chains in combination with the E2 enzymes UBE2S and UbcH10[65,66,67]. In 2010 two different crystal structures of compact Lys11 linked diubiquitin were reported. One of the structures showed Leu8 and Val70 at the interface of the two ubiquitin moieties and the two hydrophobic patches merging to form a larger surface[33]. The other structure identified a previously unreported interface on the ubiquitin helix[34]. It is now believed that Lys11 chains are able to adopt an ensemble of structures, including the two reported crystal structures; highlighting, the dynamic nature of this linkage[68].

Initial research into the function of these chains highlights a role in the degradation of mitotic regulators, with the abundance of the chains markedly increasing when cells exit mitosis[33,69]. The change in abundance is linked directly to the activation of the APC/C during the cell cycle. The APC/C is responsible for the progression through the cell cycle by targeting cell cycle regulators for degradation as well as the assembly of Lys11 chains. The ubiquitination of APC substrates with Lys11 chains was shown to be strongly influenced by two motifs known as the D box and TEK box within the substrate[70,71]. For chain assembly both motifs have been suggested to work together, the D box holds the ubiquitin in place while the TEK box in the substrate is replaced by the ubiquitin TEK box. Proteins modified with Lys11 chains have been shown to be rapidly degraded by the proteasome during cell cycle progression[65,67,71,72].

1.3.3.4 Linear ubiquitin linkages

The N-terminal methionine can be conjugated to the C terminal glycine of the adjoining ubiquitin in a head to tail manner to form linear polyubiquitin chains. Linear chains are also assembled by the E3 linear ubiquitin chain assembly complex (LUBAC)[73]. Linear diubiquitin was originally shown to be structurally equivalent to Lys63 linked diubiquitin, adopting an open conformation with no contacts between the ubiquitin subunits. The main difference to Lys63 chains is the distinctive peptide linkage[35]. The flexibility of the linker in Lys63 linked chains is not seen in linear chains as it is restrained by the N-terminal Met residue allowing the two chains to be recognised by different UBDs such as NEMO (NF-κB essential modifier) also known as IKKγ[74,75]. NEMO binds linear diubiquitin with a different topology and about 100-fold higher affinity than it does Lys63 linked diubiquitin. Linear chains, like Lys63 chains, have been shown to be essential for NF-κB signalling[74,76,77]. NEMO is a subunit of LUBAC and is required to phosphorylate the inhibitor of NF-κB (IκB), thereby inducing its degradation and releasing NF-κB to relocate to the nucleus and regulate the expression of immune genes, a role which is shared by Lys63 chains. However, a recent structure showed linear diubiquitin in a compact but not completely closed conformation[36]. It would appear that like Lys11 chains, an ensemble of structures are utilized by linear chains dependent on the binding partner.

1.3.3.5 Non conventional linkages

Very little is known about the other linkages of polyubiquitin chains and their roles have not been elucidated *in vivo*. The structures of diubiquitin molecules connected by these linkages have yet to be solved, with the exception of Lys6 diubiquitin, which was solved in 2010[78]. Lys6 diubiquitin was assembled using a chemical approach rather than using ubiquitination enzymes. The structure revealed an asymmetric compact structure distinct from previously observed structures. The proximal ubiquitin binds to the distal ubiquitin via the canonical hydrophobic surface and a second hydrophobic patch centred on Leu71 and Ile36 acts as the hydrophobic counterpart in the distal ubiquitin[78] . Despite the recent insights into the structure of Lys6 diubiquitin, much more research needs to be conducted on the structure, function and assembly of the remaining polyubiquitin chain linkages.

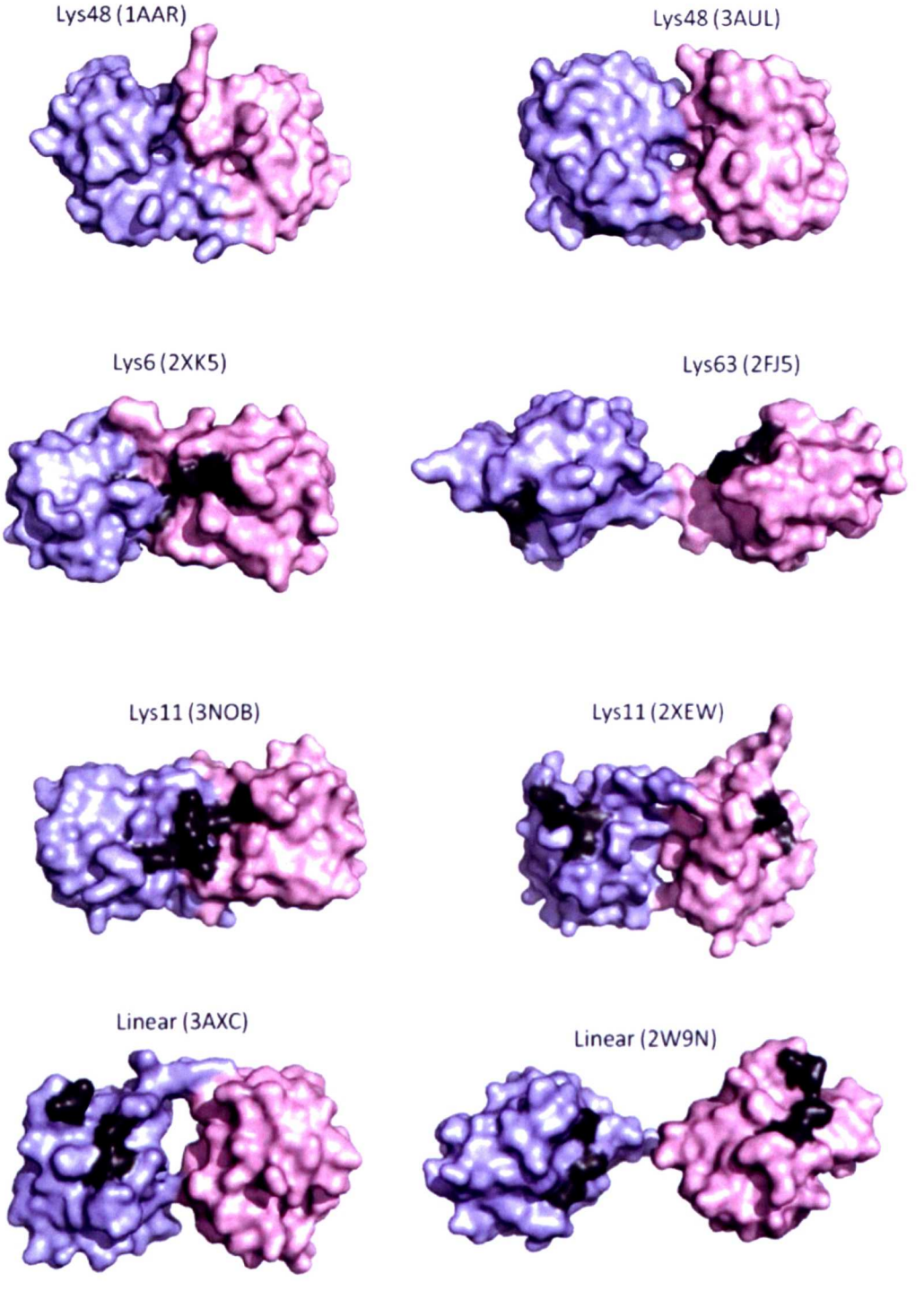


Figure 1.5 – Comparison of the structures of the different diubiquitin molecules. The PDB ID's are indicated in brackets next to each structure. The proximal ubiquitin is coloured in pink and the distal ubiquitin is coloured in blue. The hydrophobic patches (Leu8, Ile44 and Val70) from each subunit are highlighted in dark grey. From these structures it is clear that the hydrophobic patches are presented in different orientations for binding to UBDs. In the case of Lys48 (1AAR) the hydrophobic patches are not available for binding as they buried in a hydrophobic core.

1.4 Ubiqutin binding domains (UBDs)

UBDs are domains which interpret the ubiquitin code by binding non-covalently to ubiquitin. More than 20 families of UBDs have been identified to date[79,80,81]. The large number of proteins known to interact with ubiquitin strongly reflects the functional importance of ubiquitin *in vivo*. The human genome encodes over 200 proteins which contain one or more form of UBD. There are many structural differences found between the different families of UDB. Many UBDs utilize α helices for binding to the hydrophobic patch in the β-sheet of ubiquitin; however, some UBDs utilize β sheets to bind to ubiquitin and in some cases UBDs bind to other recognition sites on ubiquitin, such as Asp58. Ubiquitin-interacting motifs (UIMs), ubiquitin-associated (UBA) domains and coupling of ubiquitin conjugation to endoplasmic reticulum degradation (CUE) domains all possess α helical structures in order to bind to ubiquitin[80]. Zinc fingers are also used by UBDs but these domains tend to have roles in regulating the response to DNA damage[82]. Whilst the structure of the UBD clearly has a role in determining which type ubiquitin chains is preferentially bound, an exclusive relationship between UBD family and linkage is not observed. In a number of cases it is the spatial arrangement or the presence of multiple ubiquitin interacting sites which determines the specificity.

Structural investigations probing ubiquitin binding by different UBDs has contributed significantly to the world of structural biology. As a result, a multitude of different structures with different binding specificities have been deposited into the Protein databank (figure 1.6). For example, the Rabex-5 protein binds to monoubiquitin using the Ile44 and Asp58 binding sites using an inverted UIM and an A20 zinc finger[83], the NZF in the TAB2 protein inserts itself between the extended conformation of Lys63 linked diubiquitin[84] and the Hrs-UIM binds to two ubiquitin molecules, one on either side, both using the Ile44 surface with equal binding affinities[85]. The UBAN domain of NEMO forms a coiled coil which is capable of binding to two linear diubiquitins. Multiple contacts are made with both ubiquitins, thus conveying the specificity for linear chains[74,77]. The ZnF4 A20 binds to three ubiquitins or Lys63 linked triubiquitin. The interaction site on ubiquitin is different for each ubiquitin, with the hydrophobic patch only utilized in binding to one ubiquitin[86].

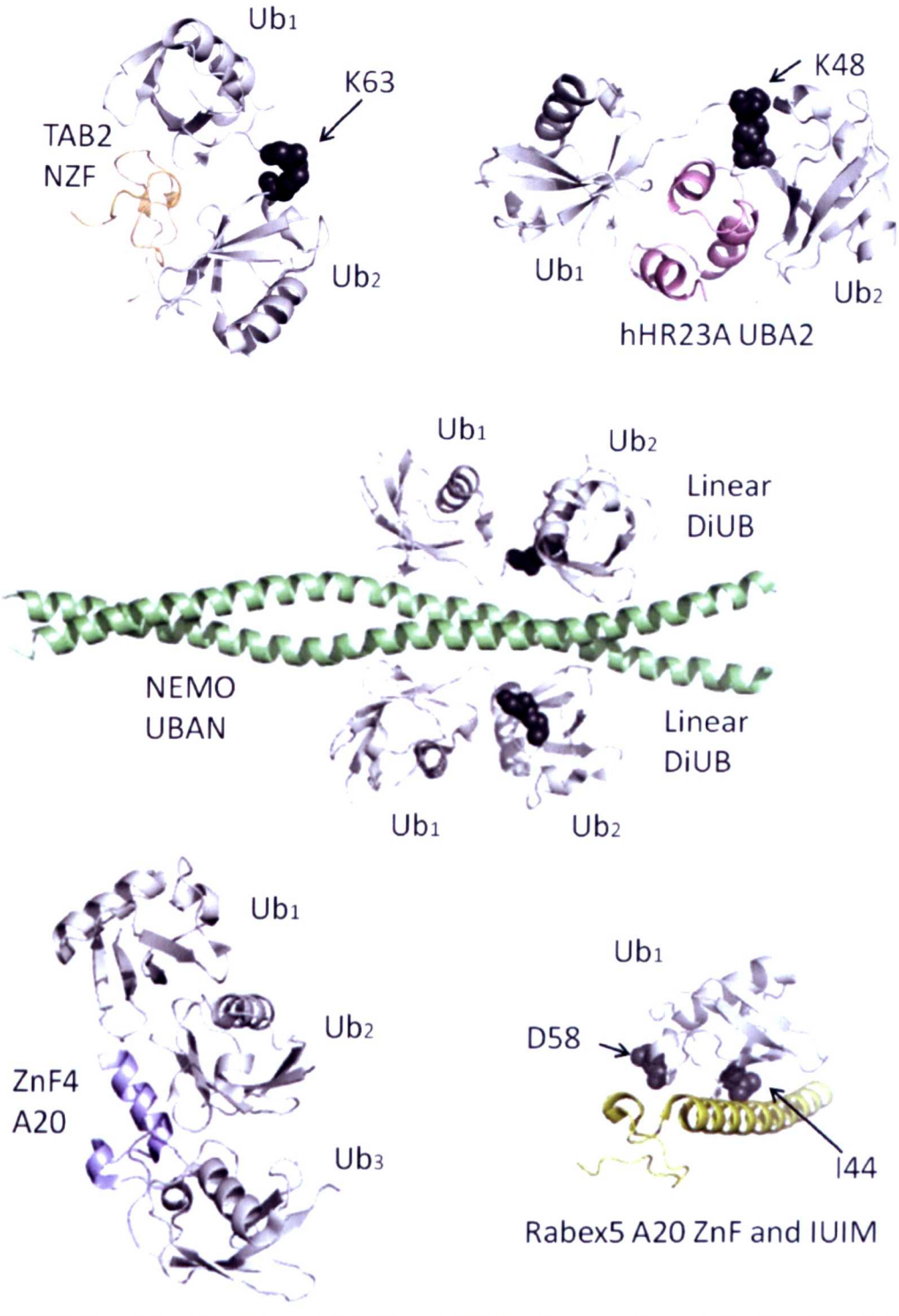


Figure 1.6 – The structures of different UBDs in complex with ubiquitin. The UBDs are coloured and the ubiquitins are grey. The TAB2 NZF domain inserts itself between the extended conformation of Lys63 diubiquitin (PDB ID 3A9J). The hHR23A UBA2 domain sandwiches between the two ubiquitins of Lys48 diubiquitin (PDB ID 1Z06). The UBAN domain of NEMO forms a coiled coil which binds to two linear diubiquitins (PDB ID 2ZV0). The ZnF4 A20 binds to three ubiquitins (PDB ID 3OJ3). The Rabex 5 binds to ubiquitin at two sites, the IUIM binds to Ile44 and the A20 ZnF binds to Asp58 (PDB ID 2FIF). The residues which link the diubiquitin moieties are shown as dark grey spheres.

1.4.1 Ubiquitin recognition by UBDs

The recognition of ubiquitin by different UBDs is extremely important. Ubiquitin recognition by single-domain proteins has been studied thoroughly[87], although the recognition of polyubiquitin chains by multi-domain proteins has only recently begun to be explored. Binding to ubiquitin using both conformational selection and induced fit mechanisms have been suggested[13,17]. Many other factors also contribute to ubiquitin recognition including, the accessibility and orientation of the hydrophobic patch in different polyubiquitin chains[5], the flexibility of polyubiquitin chains[88], polyubiquitin chain length[89], the structure of the UBD[80] and the ability of UBDs to bind to non-canonical surfaces on ubiquitin[81]. A combination of these factors enables the UBDs to interpret the different signals of the ubiquitin code.

Ubiquitin is able to adopt different conformations dependent on its binding partner. One study highlighted that for each bound ubiquitin structure there is an equivalent unbound structure, supporting the conformational selection theory[13]. Another study showed that the region surrounding the binding site on ubiquitin undergoes more pronounced conformational changes compared to the rest of the molecule, suggesting induced fit optimisations are made after binding. A molecular dynamics simulation study also hinted at interplay between the two mechanisms[90]. During a binding event, a scaffold structurally similar to the bound conformation is chosen by conformational selection, while the binding interface is subsequently optimized for specific interactions via induced fit[17].

In polyubiquitin chains, the linkage determines the positioning of the ubiquitin moieties for their recognition by UBDs[5]. The same UBD can interact with different types of polyubiquitin chain, but exhibit a preference for a specific type of chain linkage[5, 91]. As well as differences in the affinity of UBDs for different linkages, differences in affinity are also observed between polyubiquitin chains of different lengths. The affinity of most UBD-ubiquitin interactions is modest, but high affinity can be achieved using avid binding to polyubiquitin chains[92,93]. Chains must be long enough to provide sufficient binding affinity through avidity. However, chains that are too long may bind so tightly that they prevent dissociation when required[89]. Multi-domain proteins must therefore bind to chains of sufficient length to amplify the signal, yet short enough to permit dissociation

of the complex. The structure of the UBD is therefore critical in order to discriminate between the different ubiquitin signals and optimise binding to certain types of chains[80].

When the structures of polyubiquitin chains in isolation and in complex with different UBDS are explored discrepancies between them were observed. These differences highlight the flexibility within polyubiquitin chains. A recent investigation using Lys48, Lys63 and linear FRET labelled diubiquitin showed that UBDs were able to select pre-existing conformations. Certain conformations of open chains, such as Lys63 and linear chains, are able to be selected by UBDs; whereas, closed chains, such as Lys48 chains, are opened to conformations which can be recognised[88]. This study showed that a combination of conformational selection and chain remodeling are key factors in ubiquitin recognition.

Most UBDs bind to the hydrophobic patch on ubiquitin, although some bind to other regions on ubiquitin[81]. Most of the other binding sites are found close to the hydrophobic path, permitting multi-domain proteins to engage multiple binding sites simultaneously. The family and therefore the structure of the UBD often determines which region on ubiquitin is recognised[81]. For example UIM domains interact with the β5 strand, UBA domains bind to the hydrophobic patch and DUBs bind to the C-terminal tail of ubiquitin[94]. When individual interactions are probed further, different residues at the binding interface form distinct contacts with different UBDs. Ile44, the centre of the hydrophobic patch is often critical for an interaction, although in some cases it has a minor role in the affinity of the interaction[15].

1.4.2 Ubiquitin-associated domains (UBA domains)

UBA domains were the first type of motif discovered to bind to ubiquitin. They, as many other UBDs, are α-helical structures. UBA domains fold to form a compact three helix bundle. The structure is arranged to form a large hydrophobic patch around the first flexible loop region which links helices 1 and 2[95,96,97, 98]. These domains possess very little sequence homology but are structurally very similar (figure 1.7). An MGF motif in loop 1, a di-Leucine motif at the end of helix 3 and the hydrophobic patch are features common to many different UBA domains. CUE domains also fold to form three helix bundles, making these two families the most similar of the UBD families.

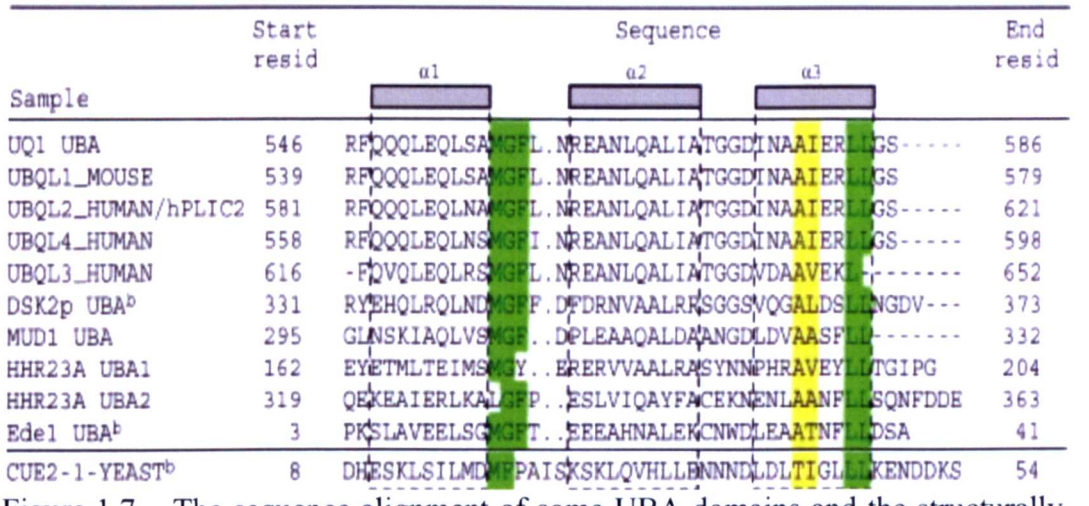


Figure 1.7 – The sequence alignment of some UBA domains and the structurally related CUE domain in the Cue2-1 protein in yeast. The helices and conserved regions are indicated. The MGF motif in loop 1and the di-Leu in helix 3 are highlighted in green. Taken from Zhang *et al*[99].

Crystal and solution structures of UBAs alone and in complex with ubiquitin are available for a variety of proteins. The complex structures reveal that the hydrophobic patch on the UBA binds to the hydrophobic patch centred on Ile44 on ubiquitin. Examples include one of the human homologues of the yeast Rad23 protein (hHR23A)[95], Dsk1[100], Mud1[101], Ede1[97] and E3 ubiquitin ligase EDD[102]. Structures identified to date highlight residues in loop 1 and helix 3 at the binding interface, suggesting a conserved interaction surface is used by UBA domains to bind to ubiquitin (figure 1.8).

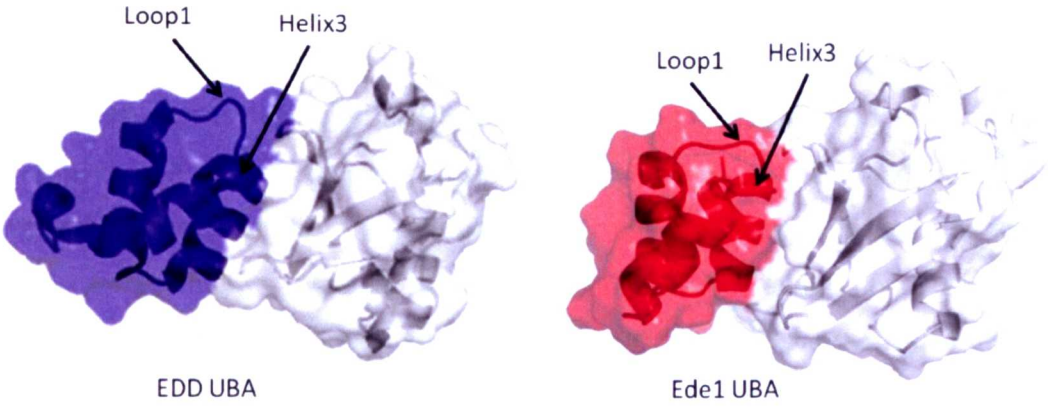


Figure 1.8 – The structures of the EDD UBA (PDB ID 2QHQ) and Ede1 UBA (PDB ID 2G3Q) in complex with ubiquitin. The three helix bundle of the UBA is coloured blue or red for the EDD UBA and Ede1 UBA respectively. Ubiquitin is coloured light grey. UBA domains bind to ubiquitin using residues in loop 1 and helix 3 highlighting a conserved interaction surface is used by UBA domains.

UBA domains bind to both monoubiquitin and polyubiquitin chains, with many UBAs preferentially binding to polyubiquitin chains[103]. Despite binding to the hydrophobic patch on ubiquitin and utilizing a conserved binding interface, a preference for a particular linkage is not displayed by UBA domains[104]. UBAs domains bind to polyubiquitin chains of different length and linkage with different affinities. For example, the hHR23A UBA2 domain exhibits a preference for Lys48 chains over both Lys63 chains and monoubiquitin[104,54]. It binds to Lys48 diubiquitin in a sandwich-like mode where the UBA inserts itself between the two adjacent ubiquitin molecules enabling it to simultaneously contact both units (figure 1.6)[54]. In contrast, the two UBA domains of Drm2 bind to Lys63 chains with higher affinity than Lys48[104]. Some UBAs, such as Ubiquillin-1 UBA, show no particular preference for chain linkage. This UBA binds to both Lys48 and Lys63 chains in a similar manner with comparable affinities[99].

1.4.3 Deubiquitinases (DUBs)

Protein ubiquitination is a reversible modification; therefore, a specific family of enzymes known as deubiquitinases (DUBs), have evolved to oppose the action of E3 ligases and remove ubiquitin from the target protein. DUBs bind to the C-terminal tail on ubiquitin and hydrolyse the isopeptide bond between ubiquitin and target proteins as well as isopeptide bonds between ubiquitin moieties in polyubiquitin chains. Approximately eighty five active DUBs are encoded by the human genome[105]. DUBs can be divided into five structurally distinct classes, ubiquitin C-teminal hydrolases (UCHs), ubiquitin specific proteases (USPs), ovarian tumour proteases (OTUs), Josephins and MNP+ metalloproteases[94,106]. The largest of the DUB families are the USPs, with 56 members. Despite having to substitute ubiquitin for an analogue, the structures of DUBs in complex with the ubiquitin analogue are available for many DUBs (figure 1.9).

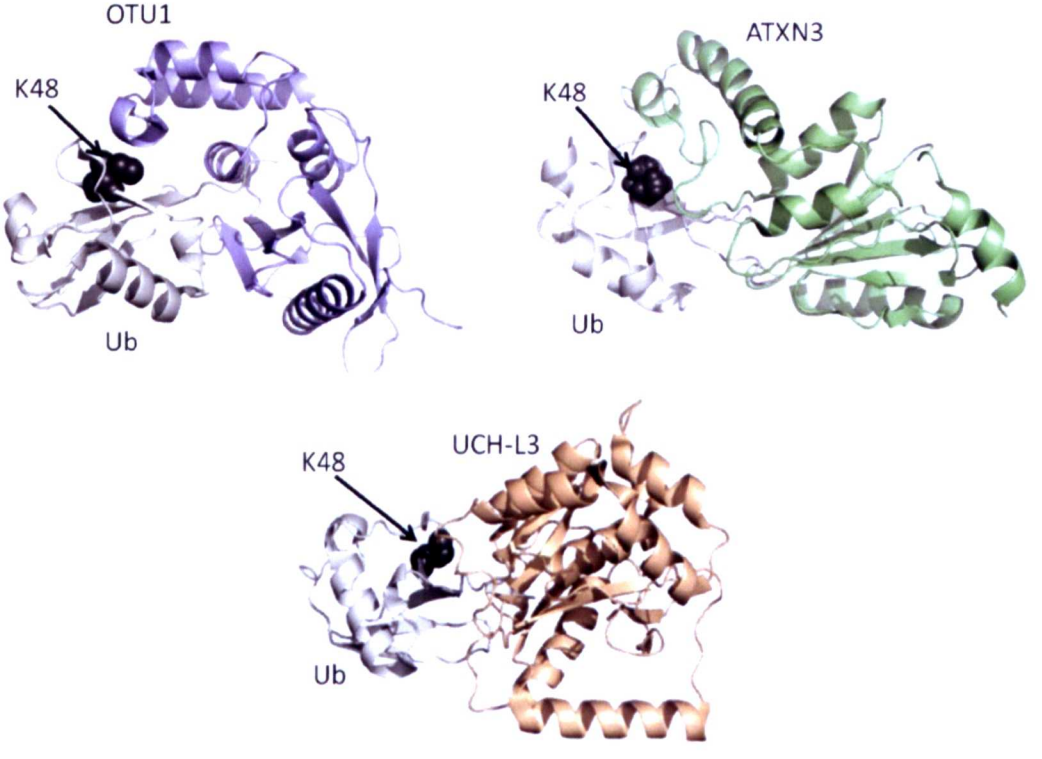


Figure 1.9 – The structures of some DUBs in complex with ubiquitin. A member of the OTU family, OTU1 in complex with ubiquitin (3BY4) is shown on the top left. A member of the Josephin family, ATXN3 in complex with ubiquitin is shown op the top right. A member of the UCH family, UCH-L3 in complex with ubiquitin is shown at the bottom. Ubiquitin in all cases is shown in light grey. Ile44 is shown as dark grey spheres.

All families, with the exception of MNP+ metalloproteases, are Cys proteases which rely of a catalytic diad or triad of residues where a nearby His side chain lowers the *pKa* of the catalytic Cys residue to permit the nucleophilic attack on the desired isopeptide bond, The third residue, usually an Asn or Asp, for those with a catalytic triad, polarizes the catalytic His[107]. Rather surprisingly the catalytic residues of all Cys protease DUBs when superimposed show only very small deviations from each other despite the families displaying vastly divergent structures[107].

Insights into the molecular recognition of ubiquitin by DUBs have been shown by biochemical studies. The C-terminus of ubiquitin extends to the ubiquitin binding site in the catalytic domain of the DUB where it is held in place for catalysis. The DUB recognizes Arg74 and Gly75 as critical residues in ubiquitin[108]. Many DUBs have been shown to possess linkage specificity, but like other UBDs the specificity is not restricted to certain families. The proteasome-assisted USP14 shows a preference for Lys48 chains[109] , whereas, CYLD has been shown to bind Lys63 and linear chains with higher affinity[110]. Some DUBs have also been shown to disassemble unanchored polyubiquitin chains, a process which involves slightly different recognition of ubiquitin. The extensively studied iospeptidase T (IsoT) is one such DUB. It utilizes a ZnF domain to recognize the C-terminal Gly-Gly motif of ubiquitin to ensure that only unanchored chains are disassembled[111].

1.5 SQSTM1/p62 protein

Over the past few decades protein communicators termed scaffold proteins have been discovered. Scaffold proteins are proteins which function to bring together interacting proteins but have no catalytic activity or specific function of their own[4]. One such scaffold protein to be identified is sequestosome1 (SQSTM1), or p62 as it is more commonly known. The p62 protein is encoded by the *SQSTM1* gene and was originally identified as binding partner of the atypical protein kinase C (aPKC)[112]. This thesis focuses on the role of p62 in NF-κB signalling[113]. p62 is capable of binding to a multitude of different proteins, including ubiquitin, to regulate a variety of cellular processes including, cell survival and cell death[114], shuttling proteins to the proteasome for degradation[115], selective macroautphagy[116,117], bone homeostasis[118,119] and NF-κB activation[120,121]. *In vivo* the p62 is often found in sequestosomes which contain large numbers of PB1 mediated oligomers and ubiquitinated proteins[122,123].

At the N-terminus lies an oligomerisation domain and at the C-terminus a UBA domain resides (figure 1.10). Between the two termini lies a zinc finger domain and a flexible, unstructured region containing several protein binding motifs. The domains of the p62 and the NF-κB signalling pathway are discussed further.

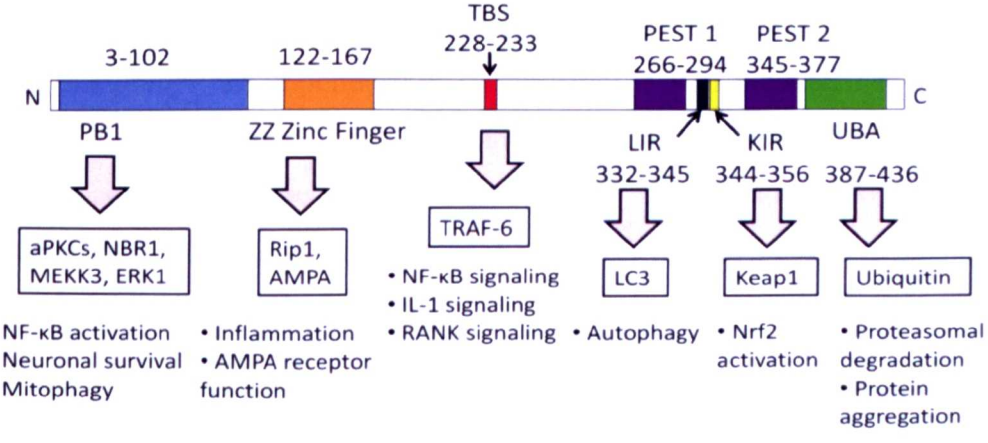


Figure 1.10– A schematic diagram of the p62 protein, its binding partners and functions. The 3 folded domains, the PB1, ZZ and UBA domains occupy residues 3-102, 122-167 and 387-436 respectively. Protein binding sites are located in the unstructured region between the ZZ domain and UBA: the TRAF-6 binding site (TBS) between residues 228-233, the LC3-interacting region (LIR) between residues 332-345 and the Keap1-interacting region (KIR) between residues 344 and 356. The 2 PEST domains, occupying residues 266-294 and 345-377, are also indicated. Some of the proteins which bind to the different regions of the p62 are shown in boxes and the pathway the interaction with p62 affects is shown below.

1.5.1PhoX and Bem1p (PB1) domain

The N terminal domain (residues 3 - 102) of the p62 protein is an oligomerisation domain known as a PhoX and Bem1p (PB1) domain. PB1 domains fold into a UbL scaffold with two α helices flanking a five stranded mixed β sheet. PB1 domains are divided into 3 different classes, type I, type II and type I/II, based on their oligomerisation capabilities. Type I contain a conserved highly acidic motif of around twenty eight amino acids known as the OPCA motif. Type II contains a conserved Lys residue capable of binding the OPCA motif on the face opposite. Type I/II contain both conserved features and can bind in a back to front manner to other PB1 domains (figure 1.11)[124]. Type I/II are therefore also capable of self oligomerisation which is inherently problematic in *in vitro* structural studies, often with insoluble aggregates formed.

The p62 PB1 domain is a type I/II PB1 which has been problematic until recently. The NMR structures of the p62 PB1 monomer[125] and dimer[126] have now been elucidated by introducing mutations to the conserved OPCA motif. The solution structures reveal two acidic clusters in the OPCA interface which are denoted A1 and A2 and are common to type I and type I/II PB1's. The opposite face reveals 2 basic clusters referred to B1 and B2'. These basic regions vary in size amongst type II and type I/II PB1. The aPKC is also classed as a type I/II PB1 domain but is unable to self-oligomerise due its small basic regions. These distinctly charged surfaces highlight the importance of electrostatics in PB1 binding. The mutations introduced to the p62 PB1 monomer and dimer are able retain the electrostatic potential of the interfaces to permit binding[125,126]. Another PB1 containing scaffold protein known as Next to BRCA1 (NBR1) was found to have the highest degree of homology to the p62 PB1[127].

The p62 PB1 domain functions to bind to other PB1 domains, including NBR1, aPKC and MEKK3. The p62 PB1 domain is also able to bind to the proteasome[128]. There has been some speculation that the PB1 and UBA domains of the p62 interact with each other as observed by other UbL-UBA proteins[129,130]. The PB1 domain is structurally distinct from a UbL domain, but it does possess a ubiquitin fold therefore there is potential for an interaction; however, no such interaction has been observed to date[131].

Figure 1.11– The ubiquitin-like structure of the p62 PB1 domain (PDB ID 2KKC). The side chain for Lys7 (the conserved Lys) is depicted in ball and stick format.

1.5.2 Zinc finger domain (ZZ domain)

The next binding motif from the N terminus is a Zinc finger domain (ZZ domain). Comprised of residues 122-167 this particular type of zinc binding domain was first identified in dystrophin[132]. ZZ domains have two co-ordinating zinc ions which are arranged in a tetrahedral manner *via* conserved Cys and His residues at opposing ends of a small β sheet. These domains are found in many scaffold proteins linked to ubiquitination and even form the basis of one of the four classes of ZZ domains. The p62 ZZ domain is an example of a scaffold ubiquitination class of ZZ domain. The p62 ZZ domain functions to bind to a serine threonine kinase known as receptor interacting protein (RIP). The interaction occurs through RIP's intermediary domain. The association between the ZZ domain and RIP links the aPKC pathway to NF-κB activation *via* TNF-α stimulation, a pathway that RIP is an intricate part of[133].

1.5.3 Protein binding motifs

With a role as a scaffold protein it is not surprising that several different protein motifs are contained within the p62 structure. For example the p62 protein was found to be a common component of inclusion bodies associated with diseases linked to protein aggregation[122]. Further study revealed a link to the degradation pathway autophagy[134], whereby a small membrane bound sac known as an autophagosome, engulfs cytoplasmic matter. A protein known as microtubule-associated protein light chain 3 (MAP-LC3) is essential for autophagosome formation. The p62 protein binds to the MAP-LC3 protein via a twenty two amino acid motif known as LC3 interacting region (LIR) which occupies residues 332 -

245[116]. The LIR therefore targets the p62 and its associated cargo for degradation by autophagy.

Immediately after the LIR lies a newly discovered motif. This motif is known as the Keap1-interacting region (KIR) because it interacts with a protein called Keap1. The KIR occupies residues 344-356. Keap1 is an adapter protein which regulates the activation of a transcription factor called Nrf2. Nrf2 is able to regulate the expression of various detoxification enzymes.

Another important motif is the TNFR-associated factor 6 (TRAF6) binding motif. This short motif is located between residues 228-233. TRAF6 is an E3 which synthesises Lys63 linked polyubiquitin chains and is a crucial element of the NF-κB activation pathway. Stimulation of TRAF6 autoubiquitination via Lys63 is mediated by the interaction with p62. p62 binds to TRAF6 via the TRAF6 binding motif is the interacting point of the p62 protein[135]. Recruitment of the TRAF6 protein to a protein/receptor complex activates the downstream pathway to directly regulate osteoclast maturation[136].

Two regions rich in Pro (P), Glu (E), Ser (S), and Thr (T) known as PEST motifs lie close to the C-terminal of p62. The PEST sequences occupy residues 266-294 and 345-377 and have been shown to be completely unstructured. Although these are not sites for p62 interaction with other proteins, the function of these sequences is thought to be linked to protein degradation. The PEST sequence is contained within many proteins with short half-lives and it is thought that the PEST sequence itself is a degradation signal[137,138].

1.5.5 Ubiquitin-associated (UBA) domain

The C-terminal UBA domain occupies residues 387 to 436. The solution NMR structures of p62 UBA domain[139] and the domain in complex with ubiquitin[140] have been solved by the Searle group. It has been shown more recently by both NMR and X-ray crystallography that the p62 UBA exists as a highly stable C2 symmetric dimer[131, 141]. This feature makes the p62 UBA different to most other UBA domains, although there are a few examples of UBA dimers (discussed in chapter 6). The somewhat unusual formation of a highly stable UBA dimer was shown to tightly regulate ubiquitin binding due to the ubiquitin binding patch being partially occluded by the dimerisation interface by the C terminus of helix

3. The dimerisation interface is formed along helix 2 of the three helix bundle (figure 1.12). An antiparellel dimer with symmetry along the C2 axis was revealed[141].

The p62 UBA domain functions to bind to ubiquitin. The p62 UBA has been shown to bind both monoubiquitin and polyubiquitin chains[140]. Although a preference for chain linkage in polyubiquitin chains has not been observed *in vitro*[142], it is expected that a preference for both Lys63 and linear chains is clear in vivo due to the involvement with the NF-κB signalling pathway. In addition to a functional role the UBA has also been shown to have a protective role by preventing the p62 from being degraded by the proteasome[143]. This is similar to other proteins which contain a C-terminal UBA domain[144].

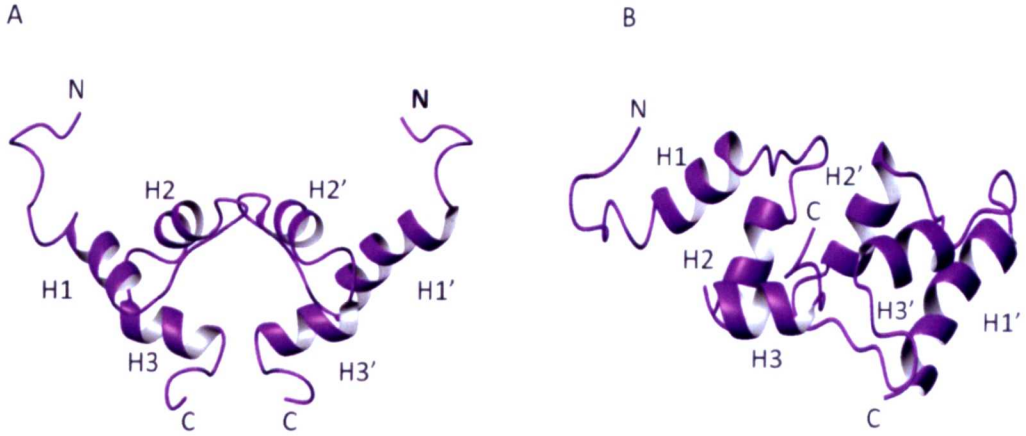


Figure 1.12– The model of the p62 UBA dimer (PDB ID 2KNV). A) the view along the C2 axis and B) rotated by 90 °C . Helix 2 and some residues in helix 3 form the dimerisation interface.

1.5.6 The p62 protein's role in NF-κB signalling

Although p62 has a role in several signalling pathways, this project is focused on the role p62 has in NF-κB signalling. NF-κB is a highly conserved transcription factor involved in activating genes responsible the differentiation and maturation of osteoclasts which are vital to maintain bone architecture. NF-κB proteins can be divided into two distinct groups, one group are synthesised as transcriptionally active proteins, whereas the other group exist as precursor molecules which require proteolytic processing to make them biologically active. Inactive precursors are bound to an inhibitor known as IkB, whereas active NF-κB proteins contain a IkB-like domain which prevents the binding of IkB[145,146].

The p62 protein is involved in the receptor activated NF-κB (RANK) activation pathway of NF-κB, which involves releasing the IkB protein from inactive NF-κB (figure 1.13). A complex is formed when RANK binds to a protein known as Rank ligand (RANKL)[147]. The RANK/RANKL complex then stimulates the recruitment of TRAF6, which is responsible for the formation of Lys63 polyubiquitin chains and is crucial for RANKL signalling[136]. The p62 protein contains a TRAF6 binding site. The p62 protein appears to initiate downstream signalling by recruiting the aPKC to the complex via interaction with its PB1 domain. The formation of the TRAF6-p62-aPKC ternary complex then stimulates TRAF6 to ubiquitinate itself via interaction with the p62 UBA domain, leading to the formation of a Lys63 linked chain[148,113]. TAB2 and TAK1 are then recruited to the complex to form a large multimeric complex. TAB2 phosphorylates TAK1 which causes Lys63 ubiquitination of NEMO[149,150]. NEMO then activates phosphorylation of IkB by IkB kinase (IKK). The resultant conformational change in IkB triggers Lys48 ubiquitination and subsequent degradation of IkB. NF-κB is then free to translocate to the nucleus where it regulates expression of genes associated with osteoclastogenesis[148,151].

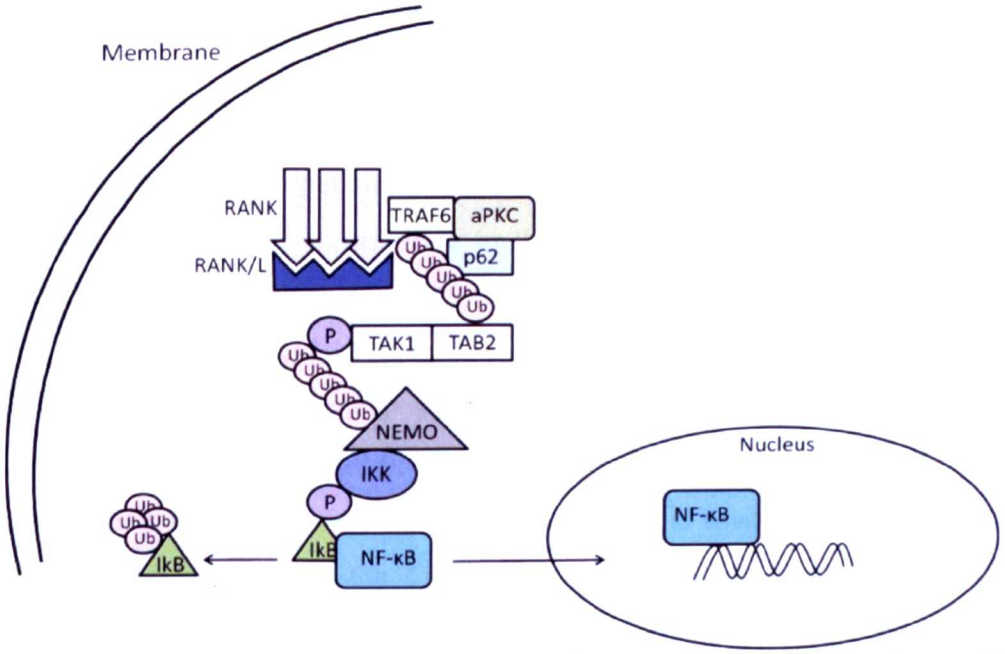


Figure 1.13 – A schematic diagram of the RANK induced pathway leading to NF-κB activation. RANKL binds to RANK to recruit TRAF6 and signal downstream activity. Lys63 ubiquitination is represented by the extended ubiquitin chains and Lys48 ubiquitination by compact ubiquitin chains. Phosphorylation events are depicted by P. The NF-κB is released from the IkB inhibitor rendering it free to translocate to the nucleus where is regulates gene expression. Adapted from Crockett *et al*[151].

1.5.7 p62 UBA and Paget's disease of bone (PDB)

Paget's disease of bone (PDB) was first described by Sir James Paget in 1876. The disorder is primarily characterised by excessive bone turnover and like many osteolytic diseases is associated with excessive osteoclastic activity[152]. A secondary increase in osteoblast activity combined with increased numbers of osteoclasts causes pagetic lesions in bones, which accounts for many of the symptoms associated with PDB. These lesions are the result of abnormal bone architecture. PDB affects 1-3% of white adults over the age of 55 in Britain[153] and although there isn't a clear pathogenesis, common symptoms include skeletal deformity, bone pain, susceptibility to pathological fractures and osteoarthritis, all of which are more serious problems in elderly patients. Mild side effects such as deafness can also occur. In many cases PDB is asymptomatic allowing polystatic progression of the disease, meaning multiple bones can become affected. A more severe complication of PDB is osteosarcoma which affects 1% of PDB patients. Although this percentage is low, the risk factor of developing osteosarcoma is 1000 times greater for PDB patients[154]. PDB is widespread in Caucasian populations. The disease is most prevalent in the UK but is also common in Western Europe and British migrants to Australia, New Zealand and South Africa[155,156].

The onset of PDB has been linked to genes encoding SQSTM1/p62 inferring a strong genetic link in Paget's disease. There is evidence for both genetic[157] and environmental factors[158] contributing to the prevalence of the disease, with the genetic side linking PDB to the p62 protein. Genetic factors have been linked to the onset of both familial and sporadic forms of PDB. Numerous studies over the past 20 years have revealed several susceptible loci, denoted PDB1 to PDB7[157,159,160] which could all be linked to a genetic predisposition of PDB. The locus mapped to chromosome 5q35-QTER (PDB3) is associated with the gene SQSTM1 which codes for the scaffold protein p62. The locus was explored further to reveal several mutations which lie within or in close proximity to the UBA domain of the p62 protein.

Initially a point mutation at position 392, P392L, was described[161]. Other PDB mutants include P387L, G411S, M404V, G425R, S399P, D423X, G425E, Y383X, A427D and I424S[162,163,164,165,166 ,167,168]. To date over thirty separate PDB mutations have been identified. However mutations outside the UBA domain, including K378X[169], have also been described. The consequences of these non-UBA mutations have yet to be established. Broadly speaking the PDB mutations can be classed as either missense or truncating mutations. Multiple studies have indicated that a loss of ubiquitin binding, which is amplified in the binding of longer polyubiquitin chains occurs for PBD mutants[170,171]. Interestingly, most of the PDB mutations are located outside of the hydrophobic patch implicated in ubiquitin binding (figure 1.14), suggesting that ubiquitin recognition by residues outside of this binding interface also have a key involvement in regulating ubiquitin binding[172]. The discovery of the UBA dimer adds further complexity to understanding the molecular mechanism which regulate ubiquitin binding.

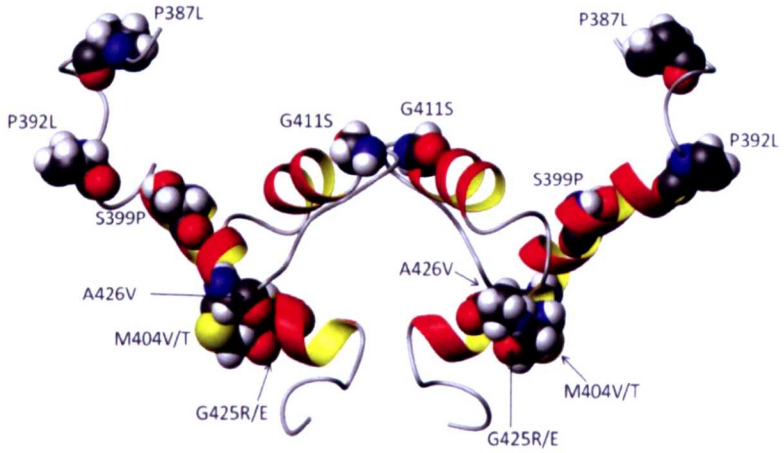


Figure 1.14 – The positions of some of the more prevalent Paget's disease mutations on the p62 UBA dimer (PDB ID 2KNV). The mutations are shown on both subunits of the UBA dimer. The mutations are not confined to the ubiquitin binding patch, suggesting other factors are involved in regulating the affinity for ubiquitin.

Substantial efforts into understanding the structural, biophysical and biological effects of the PDB mutants has been conducted as part of a collaborative project by the Searle and Layfield groups (University of Nottingham). The PDB missense mutants P392L, S399P, M404V, M404T, G411S and G425R have all been characterised structurally and biophysically. The studies highlighted no single mechanism to account for the loss of ubiquitin binding associated with PDB

mutations. Evidence for loss of structural integrity, stability of the dimer and changes to the ubiquitin binding surface have been exhibited by the PDB mutants, with a combination of the effects likely to factor into the resultant affinity for ubiquitin[143,173].

Changes in the affinity of the P392L mutant were minimal, consistent with its location away from both the dimerisation and binding interfaces. The prevalent P392L mutation is located at the N-terminus of helix 1 and does not form part of the hydrophobic patch or the dimerisation interface. This particular mutation involves structural modifications extending the N-terminus of helix 1 and moving the cap of the helix back towards residue Pro388[139]. The S399P showed a decreased affinity as a result of a disruption to the structure of helix 1. Conversely, the G425R mutant appeared to enhance the stability of the dimer, leading to a decrease in the population of monomer at equilibrium, which in turn lowers the affinity of the interaction. However, the M404V and M404T mutations cause the biggest decrease in stability and are associated with a drastic reduction in binding affinity. This result was not surprising as the Met404 residue forms part of the conserved MFG motif in loop 1 which binds to ubiquitin[173].

Biological binding studies have using the PDB mutants have also been completed since the p62 UBA has been shown to exert a preference for binding to polyubiquitin chains[170,171]. Moreover, the p62 UBA has been shown to be non-selective between Lys48 and Lys63 linked chains[142]. Pulldown binding studies have been able to explore the binding of polyubiquitin chains in detail[140] The binding of wild type p62 UBA and some of the PDB mutants (in the isolated domain) for Lys48 and Lys63 linked polyubiquitin chains is shown in figure 1.15. A slight loss of binding exhibited by the PDB mutants is also evident from these studies, although the trend of binding to polyubiquitin chains over monoubiquitin was also still present. The preference for tri and tetraubiquitin chains by both the wild type and the PDB mutants is likely reflect a preference for chains of these lengths *in vivo*. Again the M404V, M404T and G425R mutations seem to exert the biggest effects, as no evidence for binding is observed by these mutants. The molecular basis for a lack of binding is obvious for these mutants; however, the molecular mechanisms which underlie other PDB mutations are not clear.

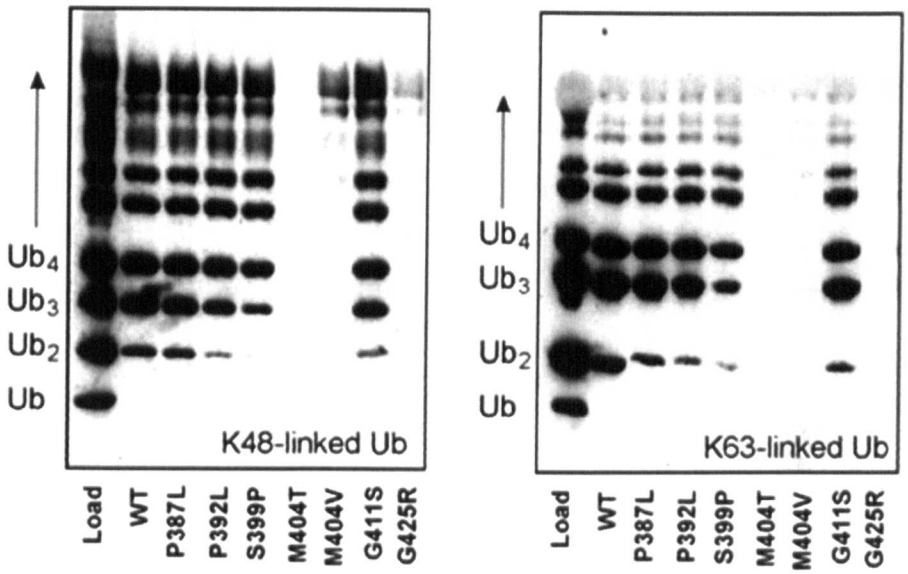


Figure 1.15 – *In vitro* binding assays which show the affinity of the isolated p62 UBA domain for polyubiquitin chains. The affinity of wild type p62 UBA and some of the more prevalent PDB mutations were investigated when bound to Lys48 or Lys63 linked polyubiquitin chains. The various UBAs were immobilized on glutathione-Sepharose beads and used to precipitate polyubiquitin chains. The precipitated proteins were then detected using Western blotting using a ubiquitin-reactive antibody. A slight reduction in affinity is observed by the PBD mutants when compared to the wild type p62 UBA. Also no preference for Lys48 or Lys63 chains is observed, but tri and tetraubiquitin chains are shown to bind with the highest affinity for both the wild type p62 UBA and the PBD mutants. Taken from Long *et al*[140].

Despite structural and biophysical characterization overall only small changes in affinity are exerted by these PDB mutants, with more pronounced effects observed in the binding of polyubiquitin chains over monoubiquitin. It is therefore predicted that the small changes in affinity are raised to high affinity interactions by oligomeric p62. Moreover, these mutants exhibited differences in affinity when inserted into the isolated domain and the full length p62 protein[170,171], suggesting that other factors are also involved in regulating ubiquitin binding by p62. The presence of the non-UBA PDB mutations supports this hypothesis. Since all of the structural investigations have occurred in the isolated UBA domain, the structure of the full length protein is a long term aim. Despite significant advances to the field, the molecular mechanisms which underlie PDB remain unknown.

1.6 Aims of this investigation

The ubiquitin modification is far more complex than any other protein PTM, owing to the large size of the covalent modification and the ability of ubiquitin to polymerise forming a variety of structurally distinct polyubiquitin chains. These polyubiquitin chains present different surfaces for interaction with UBDs. The key to 'cracking the ubiquitin code' potentially lies in the ability of UBDs to recognise polyubiquitin chains of a specific linkage and length. The mechanism of binding by UBDs has shown to be diverse despite only a limited number of structures being available.

Defective ubiquitin recognition and binding has been implicated in patients with PDB. The genetic evidence clearly points to mutations within the p62 UBA contributing to the onset of PDB. The experiments conducted in this thesis focus on ubiquitin binding by the p62 UBA. It is likely that, a combination of factors are able to act in concert to tightly regulate the affinity of the p62 UBA. Differences in the affinity of the UBA in the isolated domain and the full length protein have been observed, suggesting factors outside of the UBA are also involved in regulating affinity.

In the UBA, factors, such as phosphorylation and the discovery of the p62 UBA dimer, have been shown to influence the affinity of the p62 UBA. We aim to explore the effects of dimerisation (chapter 6) and phosphorylation (chapter 7) in ubiquitin binding using structural and biophysical techniques. Mutations at the dimer interface in the isolated p62 UBA domain (T414A, R415K, Q418A, T419A, T414K, R415K, Q418K and T419KA) were used to gain an insight into the role dimerisation has on UBA affinity. Phosphomimetic mutants (S403D and S403E) and a dephosphorylated control (S403A) were also engineered in the isolated p62 UBA domain and used to investigate the effects phosphorylation has on UBA affinity. A greater understanding into how ubiquitin binding is regulated by the p62 UBA may aid in understanding the molecular mechanisms which underlie PDB. The small changes in affinity exhibited by the PDB mutants could be amplified if the stability or structural integrity of the dimer is altered or whether the UBA is phosphorylated or not.

Full length p62 functions as a molecular scaffold which can regulate multiple pathways simultaneously. However, in order to investigate the ability of p62 to bind multiple proteins is hindered by the problems associated with purifying the full length protein. Full length p62 is difficult to study due to its highly oligomeric nature and large unstructured region. Recently it has emerged that p62 has a role in another degradation pathway known as selective autophagy. As the p62 contains both a UBA and a LIR it is capable of binding to both ubiquitin and MAP-LC3. The p62 can bind to ubiquitinated proteins and deliver them to either the proteasome or target them for degradation by autophagy. Our objective is to study the interactions with both MAP-LC3 and ubiquitin and investigating whether a p62 mediated ternary complex is able to be formed by the three proteins (chapter 8) and whether an allosteric relationship exists between the binding proteins. The p62 ternary complex was investigated using a p62 subfragment encoding residues 300-440, which encompasses the interaction site for both binding proteins. This was mutated to a C331S p62 300-440 variant in order to prevent disulphide bonds forming between molecules.

Full length p62 is found in cellular speckles due to its ability to self oligomerise via N-terminal PB1 interactions. Therefore, arrays of UBA domains are presented to polyubiquitin chains *in vivo*. The affinity of the individual UBA-ubiquitin interactions can therefore be amplified by avidity effects. The small decreases in affinity observed by PBD mutants are likely to be enhanced in patients suffering from the disease. Avidity effects could potentially explain the subtle changes in affinity observed by the highly prevalent P392L PDB mutation. The main aim is to discover how weak p62 UBA:ubiquitin interaction are 'levered' to physiologically relevant high affinity interactions (chapter 9). In order to achieve this, a model of the full length protein was designed. The model contained GST dimers to mimic PB1 domains and fragments of the p62 extending back from the C-terminus containing the UBA (GST-p62 UBA, GST-p62 341-440, GST-p62 300-440 and GST-p62 261-440). The model was used to probe differences in affinity of the UBA when binding to either monoubiquitin or linear diubiquitin.

2.0 Biophysical techniques

This thesis focuses mainly on the p62 UBA and its interaction with ubiquitin; however, longer subfragments of the p62 were expressed and purified in order to investigate the behaviour of the full length p62 and its interactions with other proteins, such as the MAP-LC3 protein. In order to probe protein structure and study protein-protein interactions a variety of techniques can be used. The techniques which were used throughout the course of this study are described further.

2.1 Electrospray ionization mass spectrometry (ESI-MS)

Mass spectrometry is a sensitive technique used to detect proteins at physiological concentrations. The two most widely used types of mass spectrometry are matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI)[174]. Molecules are ionized and transferred to the gas phase from solution in both MALDI and ESI-MS. The use of electrospray as a ionization technique has revolutionized the field of mass spectrometry as it overcomes the tendency of large molecules, such as proteins, to fragment during ionization. In recognition of the achievement, its inventor Prof John Bennett Fenn was awarded the Nobel prize in Chemistry in 2002[175]. A range of information can be deduced from ESI-MS including molecular mass, the stoichiometry of non-covalent protein complexes and the affinity of the interactions within protein complexes[176]. The soft ionization process of electrospray enables fragile protein complexes to be transferred to the gas phase intact even if the interaction is weak. Larger macromolecular complexes can be studied using nanospray ESI, a miniature version of ESI-MS[177].

A mass spectrometer consists of three mains parts, an ion source, a mass detector and a mass analyser (figure 2.1). Samples are dissolved in a polar, volatile solvent and are pumped through a narrow steel capillary. A high voltage is applied to the tip of the capillary where the liquid becomes unstable as it is forced to hold more and more charge. Eventually the liquid reaches a critical point where it can hold no more charge and Coulombic explosion occurs. The liquid is then dispensed as a fine aerosol of highly charged droplets. The droplets decrease in size as a result of solvent evaporation producing ions which are released from the droplets. A

nebulising gas, such as nitrogen, normally flows outside of the capillary and is sprayed towards the mass spectrometer. A spectrum of multiple charge states is visualised and separated according to their mass to charge ratio (m/z).

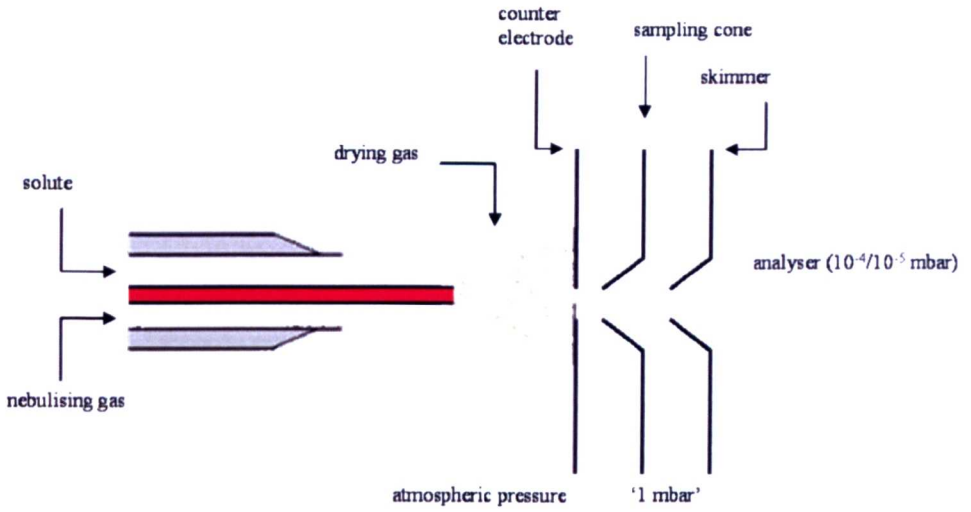


Figure 2.1 – A schematic representation of an electrospray ionisation source, showing the different regions that the sample travels through before detection in the mass analyser

Mass spectrometry has become very useful to the study of proteins and protein complexes. Its main advantage over other techniques is its ability to use very low sample concentrations which are much nearer to physiological concentrations. This is especially useful for proteins which are prone to aggregation at higher concentrations. Moreover the data collection can occur in minutes. However, there has been much debate as to whether the spectrum produced is a true representation of molecules in solution. For example non specific aggregates can form which are not present in solution. In protein-protein interactions the type of interaction can also play a major role in the spectrum that is produced. Hydrophobic interactions are weakened but not completely abolished *in vacuo* due to the absence of water which drives the hydrophobic effect[178,179]. Whereas electrostatic interactions are greatly enhanced *in vacuo* because of the absence of solvent shielding[180]. Therefore, some protein-protein interactions can be more difficult to monitor than others using ESI-MS.

2.2 Far Ultra-violet circular dichroism (Far UV-CD)

Circular dichroism (CD) spectroscopy measures the difference in absorption between left and right handed circularly polarized light. Polarized light can either be linear or circular. Linear polarized light is light which oscillates only in one plane, circularly polarized light occurs when the direction of the electric field rotates about its propagation direction (figure 2.2). When polarized light passes through an asymmetric molecule elliptically polarized light is generated, which is known as CD signal.

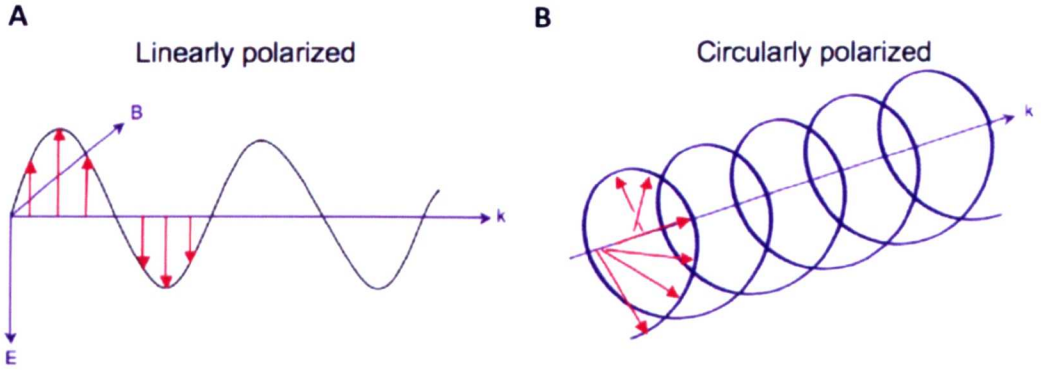


Figure 2.2 – A diagram showing the electric vectors of A) linearly polarized light and B) circularly polarized light at a moment in time. Electromagnetic radiation consists of an electric (E) and magnetic (B) field that oscillate perpendicular to one another and the propagating direction (k). Linearly polarised light occurs when the electric field vector oscillates only in one plane whereas, circularly polarized light occurs when the direction of the electric field vector rotates about its propagation direction.

Molecules which have structural asymmetry exhibit positive and negative CD signals; whereas, symmetrical structures exhibit a CD signal of zero. Proteins are optically active molecules due to their asymmetric chiral centres, meaning they emit a CD signal. Far UV-CD refers to wavelengths between 190-250 nm and is ideally suited to monitor the extent of secondary structure within proteins[181]. Near UV-CD refers to the wavelength range 250-300 nm and can reveal information on the tertiary structure of proteins.

When amide polypeptide bonds are aligned in regular arrays, such as α helices or β sheets, the different secondary structure elements produce characteristic far UV-CD spectra (figure 2.3). Proteins with disordered structures can also be readily identified by their CD spectra as they have very low ellipticity above 210 nm and

a negative band at 195 nm[182]. Proteins with a large α helical content produce negative bands at 222 nm and 208 nm and a positive band at 193 nm[183]. Whereas proteins with well defined β sheets produce a negative band at 218 nm and a positive band at 193 nm[184].

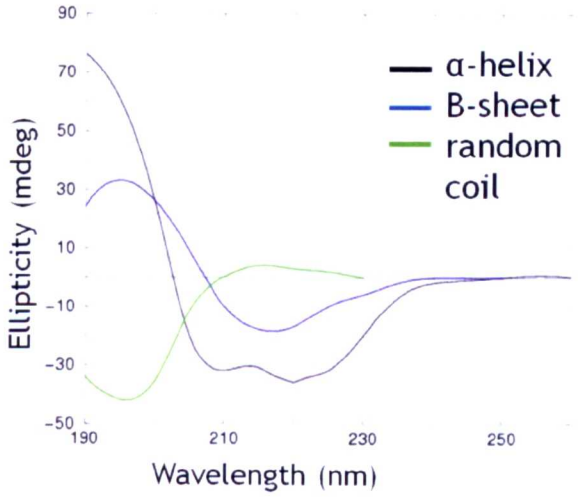


Figure 2.3 – The characteristic CD spectra for the different secondary structure elements of proteins. The α helix and β sheet are shown by the black and blue lines respectively. Proteins with a random coil structure also produce distinctive CD spectra as indicated by the green line.

Since the secondary structure elements possess such distinguishing CD spectra, the unfolding of proteins can also be monitored using Far UV-CD. Unfolding profiles can be collected as a function of temperature[185], denaturant or osmolyte[186]. The free energy of folding can be determined from CD measurements as a function of these factors if the change in CD is due to a transition between two states, the folded and unfolded state. The free energy of folding is linearly dependent on the concentration of denaturants, such as urea or guanidine, or increasing temperature which both cause a loss of ellipticity. Osmolytes, for example trimethylamine N-oxide (TMAO) or sucrose, increase the stability of proteins and cause an increase in ellipticity. These types of experiment can be particularly useful if used to determine whether the stability has been affected by certain mutations. Thermal unfolding experiments, as used in this thesis, were recorded at 222nm, as this is the wavelength where the greatest difference in CD signal is observed between α helical structures and random coils.

2.3 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a highly sensitive technique used to monitor binding events, whether the interaction is protein-protein or protein-ligand. During a binding event there is a change in the Gibbs free energy as a result of bond formation or bond breaking and the uptake or release of water molecules. Therefore various entropic and enthalpic factors contribute to the formation of a complex. Despite its more common application to monitoring association events, ITC has also been shown to be a useful technique for monitoring protein dimer dissociations[187].

The ITC calorimeter consists of two identical cells, the sample cell and a reference cell which contains water or buffer (figure 2.4). For both types of experiment a protein of known concentration is loaded into a syringe which is injected into the sample cell. For binding events the ITC experiment involves sequentially injecting one protein into another; whereas, for dissociations a dimeric protein is injected into a cell which contains a degassed buffer solution.

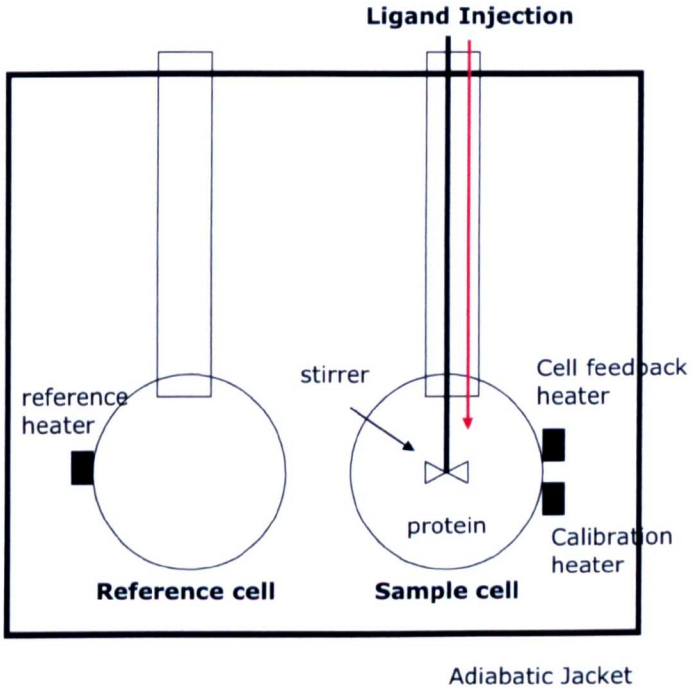


Figure 2.4 – a schematic diagram of the ITC calorimeter. The diagram shows typical set up for a binding experiment, whereby one protein is loaded in the sample cell and the ligand (or second protein) in the injector syringe. In a dissociation experiment the protein in the sample cell is replaced with a degassed buffer solution.

Both the sample cell and reference cell are made of an efficient thermal conducting material and are enclosed by an adiabatic jacket. Sensitive thermophile circuits regulated by a feedback control system are used to detect any change in temperature. These circuits also ensure that the two cells are kept at thermal equilibrium throughout the experiment. Prior to the start of an experiment, a constant power is applied to the reference cell. This directs a feedback circuit, activating a heater located on the sample cell. Temperature changes in the sample cell are compared to the reference cell after each injection in the experiment. In an exothermic reaction heat is generated which causes an increase in the temperature within the sample cell. This causes the feedback circuit to decrease the power to the sample cell. In contrast, in an endothermic reaction, heat is taken in causing the feedback circuit to increase its power to return the temperature back to thermal equilibrium with the reference cell.

The raw ITC data consists of a series of spikes of heat flow, each corresponding to an injection (figure 2.5). As the experiment continues these spikes decrease in size until they reach complete saturation. The area under each peak is then integrated with respect to time and plotted against the molar ratio of the components. The data is then fitted to the appropriate binding or dissociation model. Binding affinities (K_a) or Dissociation constants (K_d), enthalpy changes (ΔH) and the stoichiometry (n) of a reaction can all be directly determined from a single ITC experiment. Gibbs energy changes (ΔG) and entropy changes (ΔS) can then be determined indirectly to generate a full thermodynamic profile.

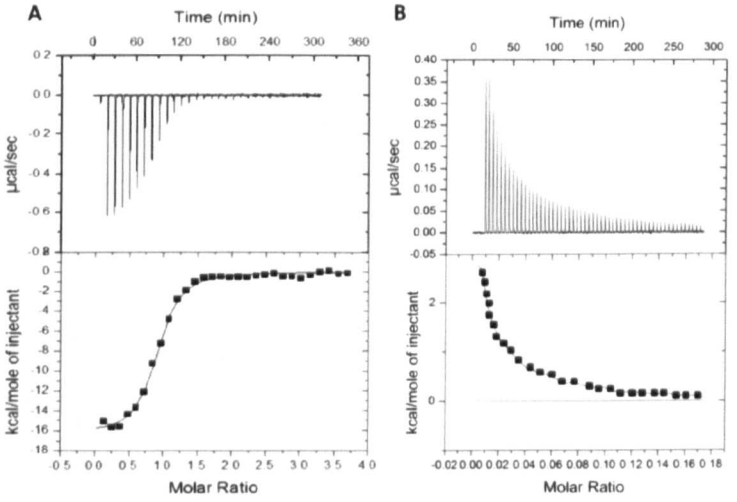


Figure 2.5 – The raw heat response generated by the ITC for A) a typical 1:1 binding interaction and (B) a typical dimer dissociation.

2.4 Nuclear magnetic resonance (NMR) of proteins

Nuclear magnetic resonance (NMR) is a powerful technique used to determine the structure of proteins. NMR produces data in atomic detail enabling a structure of the molecule to be deduced. Although large amounts of protein (up to 1 mM) and the time required to collect high quality data can be lengthy, the benefits of using NMR to study protein structure and protein-protein interactions far outweighs any of its limitations. NMR can also provide information on protein dynamics, intramolecular bonding patterns and protein folding.

NMR experiments are normally conducted in solution. Native-like structures are generated because the data is obtained using near-physiological conditions. Although, the development of solid state NMR has advanced significantly, providing substantial insights into the structures of membrane bound proteins[188,189]. The use of NMR in structure determination is usually restricted to smaller proteins; however, the use of high-field spectrometers and highly sensitive cryogenic probes have dramatically improved the quality of the data enabling larger proteins to also be studied[190]. Proteins with a highly ordered tertiary structure produces resonances with a wide dispersion of chemical shifts, whereas, intrinsically disordered proteins produce somewhat more complicated spectra due to significant overlap of resonances.

2.4.1 Monitoring the purity and folding of proteins

A one dimensional (1D) spectrum is normally recorded on an unlabelled protein sample in order to check the suitability of the protein sample for further NMR experiments. From the spectra any small molecule contaminants from the purification can be detected. These contaminants produce sharp peaks with much stronger intensity. These contaminants can sometimes be identified from the resonance ppm they appear at. The contaminants should ideally be removed before proceeding with further NMR experiments. Perhaps more importantly the 1D spectrum can also indicate whether the protein is folded. A wide dispersion of peaks is normally indicative of a folded protein. If a narrow distribution of peaks between 7.5 and 8.5 ppm are observed the protein is likely to be unfolded. Once the 1D spectrum has been recorded for a protein and it appears to be in its correct conformation then more complex NMR experiments can be conducted.

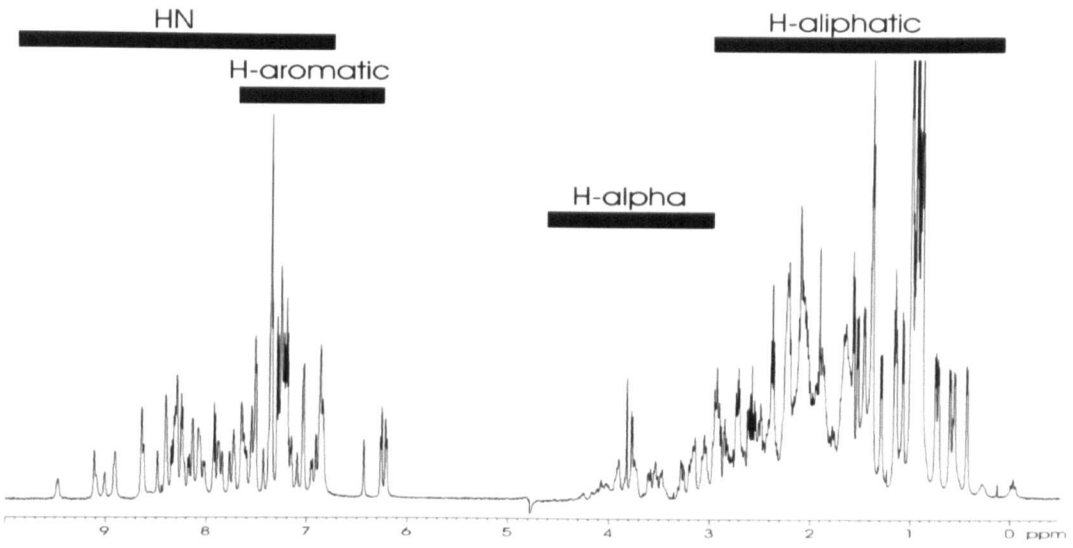


Figure 2.6 – The 1D proton spectrum of the UBA2 domain of the hHR23A protein. The characteristic frequency ranges of the various resonances are highlighted. The wide dispersion of peaks suggests that this domain is folded.

2.4.2 Monitoring protein-protein interactions using NMR

Since NMR produces atomic level data, information on each residue during a binding event can be monitored. NMR has therefore been used extensively to study protein-protein interactions. During the binding event the amino acids which interact usually undergo a conformational change which can be detected using NMR. NMR is particularly useful for probing weak protein-protein interactions due to the sensitivity of the chemical shift of the observed resonances to their local environment[191].

One of the most useful NMR experiments is the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC), which correlates the ^{1}H-^{15}N chemical shifts of amide groups[192,193]. The backbone of the protein contains an amide group for each amino acid thus in a ^{1}H-^{15}N HSQC experiment a single peak appears for all amino acids except Pro, which lacks a backbone amide group. Peaks are also observed for nitrogen-bound protons in the side chains of Arg, Lys, Gln and Asn residues. The chemical shifts of the peaks observed in a ^{1}H-^{15}N HSQC experiment are highly sensitive, with movement being observed by resonances directly involved in an interaction. Chemical shift perturbations (CSPs) are therefore considered to be a powerful tool to monitor changes upon binding to another protein, or ligand. In general the bigger the CSP, the closer to the binding pocket the residue is.

The use of the ^{1}H-^{15}N HSCQ experiment is usually limited to smaller proteins. The tumbling rate for larger proteins is slower which leads to shorter relaxation times, line broadening and poor resolution. The introduction of ^{1}H-^{15}N transverse relaxation optimized spectroscopy (TROSY) experiments as an alternative to the ^{1}H-^{15}N HSQC experiment permits the study of larger proteins. The ^{1}H-^{15}N TROSY experiment reduces the line width by selecting for the effects of dipole-dipole coupling and chemical shift anisotropy (CSA) which mutually cancel, giving rise to a single sharp peak in the spectrum[194].

Titration experiments, whereby one protein is gradually titrated into another, are often conducted using ^{1}H-^{15}N HSQC/^{1}H-^{15}N TROSY experiments. The sensitivity of the backbone amide group chemical shift to its local environment can be exploited to study structural changes in the protein. Such structural alterations can be caused by mutation, construct extension or binding events. The changes in chemical shifts between the free and bound forms of the protein can mapped to the surface of the molecule to identify binding patches. During the titration, the chemical exchange regime adopted by the protein can be established. This can often provide information on the strength of an interaction. There are three types of chemical exchange which can be utilized by proteins during a binding event, namely slow, intermediate and fast exchange (figure 2.6). Different chemical exchange processes arise from the size of the rate constant with respect to change in chemical shift. In fast exchange the rate is larger than the difference in chemical shift between the two states resulting in peaks which are a population weighted average being observed in the spectrum. This allows peaks to be tracked by movement between the free and bound forms. In slow exchange signals for both the free and bound forms are observed because the rate is less than the difference in chemical shift. The intensities of the peaks change during the course of the titration and are proportional to the populations of each species. Intermediate exchange is a somewhat more challenging regime to interpret as signals from both fast and slow exchange are observed resulting in very broad signals.

In order for titration data to be used to map the residues involved in binding onto the surface of a protein, the ^{1}H-^{15}N heteronuclear experiments need to be fully assigned. This can be achieved using a variety of different experiments, which are discussed later.

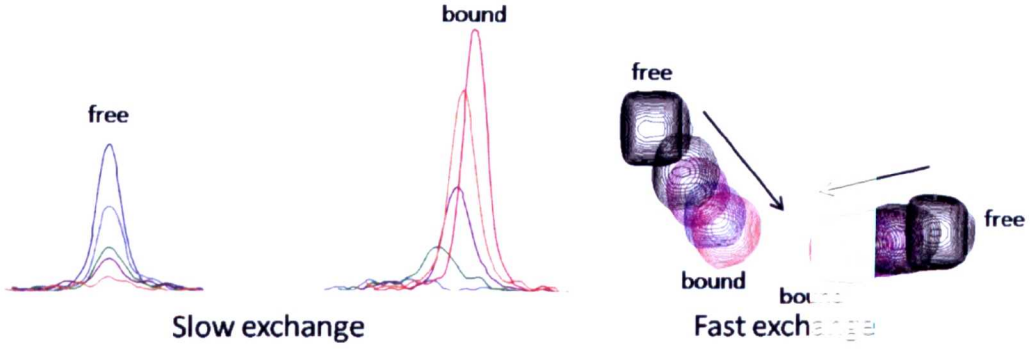


Figure 2.7 – Fast and slow chemical exchange regimes. In slow exchange (left), signals for both the free and bound forms are observed. When a binding ligand is added the intensity of the peaks corresponding to the free form decrease as the equilibrium shifts to favour the bound form. Concurrently peaks corresponding to the bound form increase in intensity. In fast exchange (right), a population weighted average is observed which moves from the free to the bound form.

2.4.3 Determining protein flexibility using NMR

The conformational flexibility of proteins is directly linked to its function. Motions in proteins are required for both substrate recognition and binding events; therefore, the ability to monitor the dynamic nature of proteins is crucial to understanding their biological function. The heteronuclear NOE (^{15}N (^{1}H) NOE) experiment is the most extensively used NMR experiment to determine protein backbone dynamics in solution[195,196,197]. Motions on a fast time scale (picoseconds – microseconds) in atomic detail can be detected using this type of experiment[198]. The ^{1}H-^{15}N NOE values are calculated by taking the ratio of ^{15}N signal intensities recorded in the presence and absence of proton saturation preceding the evolution of ^{15}N magnetization[198]. The experiments are recorded as two interleaved ^{1}H-^{15}N HSQCs. Since the intensity ratio of the two independent experiments is taken, a complete recovery of ^{15}N magnetization during the delay is critical to obtain accurate NOE values.

Structured and unstructured regions of proteins can be readily identified in heteronuclear NOE experiments. The motion of individual amide bonds relative to the tumbling rate of the overall protein can be calculated, yielding ^{1}H-^{15}N NOE values with different signs and magnitudes. Positive NOE values correspond to amide bonds which are tumbling with a correlation time comparable to the overall protein. Amino acids which produce positive NOE values are found in folded regions of the protein. Large negative NOE values represent unstructured, highly flexible regions of the protein and are produced by residues with amide bond motions that occur on a faster time scale relative to the overall tumbling rate of the protein. One major disadvantage to the heteronuclear NOE experiment is that is has inherently low sensitivity, which is approximately ten times lower than the standard ^{1}H-^{15}N HSQC. However, the use of cryoprobes in recent years has helped to overcome this problem and improve the sensitivity of the experiment. The effects of ^{15}N chemical shift anisotropy and ^{1}H-^{15}N dipole are greater when larger magnetic fields are used so this should be taken into consideration when calculating NOE values[199]. In order to obtain the maximum information from a heteronuclear NOE experiment a full assignment of the protein is required.

2.4.4 Protein backbone assignment

The backbone of a protein refers to the Hα, N_H, N_H, Cα, Cβ and C' resonances. A protein of *n* amino acids contains *n* spin systems which can be assigned to a particular residue within the protein sequence. In order to assign the protein backbone various different approaches can be made using unlabelled, ^{15}N labelled or $^{13}C/^{15}N$ labelled proteins. The natural abundance of ^{14}N and ^{12}C far outweighs that of the ^{15}N and ^{13}C isotopes; therefore, isotopic labelling of proteins is required for larger proteins. More complicated NMR experiments which require isotopic labelling tend to take longer to obtain high quality data. For unstable proteins these types of experiment are often unsuitable; however, the development of SOFAST[200] and BEST[201] pulse sequences in which increased signal-to-noise ratios are achieved has helped to overcome such problems.

Unlabelled proteins are used to conduct 2D homonuclear experiments as they only contain proton nuclei which can be detected by NMR. The homonuclear experiments commonly used for backbone assignment are 2D Nuclear Overhauser effect spectroscopy (NOESY) and 2D total correlation spectroscopy (TOCSY). In a 2D NOESY experiment magnetization is transferred from one spin system to another through space as a result of a relaxation process known as the Nuclear Overhauser effect. In a TOCSY experiment magnetization is transferred between spins which are connected to one another through chemical bonds. The TOCSY experiment is used to identify amino acid type; whereas, the NOESY experiment is used to sequentially connect the amino acid sequence. The experiments are usually run together as they complimentary to each other.

It is possible to attain a full backbone assignment using only a ^{15}N labelled protein. However, this only possible for small proteins due to the high degree of resonance overlap observed by larger proteins. The experiments used for double resonance backbone assignment are the ^{15}N HSQC-NOESY and the ^{15}N HSQC-TOCSY[202,203,204]. These experiments have one nitrogen dimension and two proton dimensions. Assignment is achieved using the same approach as for the homonuclear NOESY and TOCSY experiments.

For proteins which are bigger than 10 kDa, isotopic labelling with both ^{15}N and ^{13}C are required to the complete assignment. When proteins have been $^{15}N/^{13}C$

labelled triple resonance NMR experiments can be conducted to assign the protein backbone. These experiments have three dimensions - one proton, one nitrogen and one carbon dimension.

In order to visualise ether double or triple resonance 3D spectra strips are required. Strips are 2D segments which are narrowed in two of the three dimensions but show the total width of the third dimension. The simplest way to visualise a 3D experiment is to think of the 2D experiment which has been extended into a third dimension. The 2D experiments are normally a ^{1}H-^{15}N HSQC or ^{1}H-^{15}N TROSY for triple resonance experiments or ^{15}N HSQC-NOESY or ^{15}N HSQC-TOCSY for double resonance experiments. By selecting a peak from the 2D which is defined by a ^{1}H and ^{15}N value, a strip showing the resonances present in the third dimension can be seen.

In order to determine Cα and Cβ chemical shifts HNCACB[205] and HN(CO)CACB[206] experiments were conducted. In addition to this a HNCA experiment was recorded to ensure detection of Cα chemical shifts because the HNCACB and HN(CO)CACB experiments were optimized for Cβ chemical shifts. The HNCACB experiment correlates the amide proton and nitrogen to the Cα and Cβ for both the *i* and *i-1* residues. Therefore four peaks should be observed for any given strip, except those containing a Gly residue, which lacks a Cβ. One of the most useful features of the HNCACB experiment is that the correlations of the Cα and Cβ nuclei are in opposite phasing making them easy to identify. The HN(CO)CACB experiment is complementary to the HNCACB because it identifies the Cα and Cβ for the *i-1* residue. Therefore 2 peaks are seen for all amino acids except Gly. When the two are overlaid the Cα and Cβ chemical shits belonging to the *i* and *i-1* residue can be confirmed. The HNCA[207] [208] experiment shows the Cα for the *i* and *i-1* residues and can be used to aid identification where the Cα chemical shifts were either weakly detected or not observed. The HNCA experiment is quite useful as it is more sensitive than either the HNCACB or HN(CO)CACB experiments due to the decreased effect of carbonyl chemical shift anisotropy. In all experiments the coupling to the Cα and Cβ chemical shifts of the *i* residue is stronger meaning these peaks will appear with greater intensity in the spectra. The flow of magnetization in all these experiments is shown in figure 2.7.

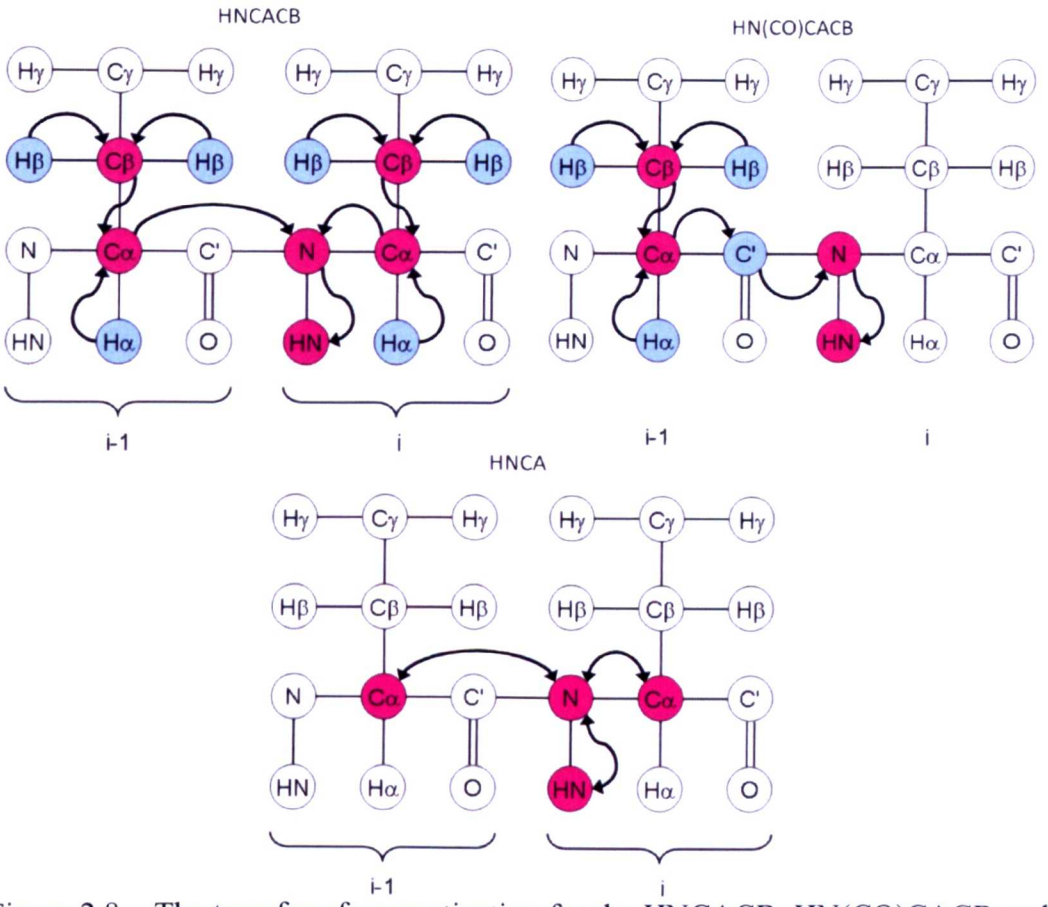


Figure 2.8 – The transfer of magnetization for the HNCACB, HN(CO)CACB and HNCA triple resonance NMR experiments which were used in this thesis. These experiments were used to identify the chemical shifts of Cα and Cβ[209].

In order to determine the carbonyl carbon (C') chemical shifts HNCO[207,208,210] and HN(CA)CO[211] experiments were conducted. The HNCO identifies the C' for the *i* residue; whereas, the HN(CA)CO identifies the C' for both the *i* and *i-1* residues. Like the HNCA, the HNCO experiment is very sensitive. The HNCO and HN(CA)CO experiments are complementary to each other much like the HNCACB and HN(CO)CACB. When overlaid the C' belonging to either the *i* or the *i-1* residue can be confirmed. Again in these experiments the C' resonances belonging to the *i* residue have a greater intensity than those observed for the *i-1* residue. These spectra are harder to use because there is still a significant degree of overlap due to a smaller frequency range in the carbon dimension. Once identified the C' shifts are linked to the Cα and Cβ chemical shifts. The flow of magnetization in all these experiments is shown in figure 2.8.

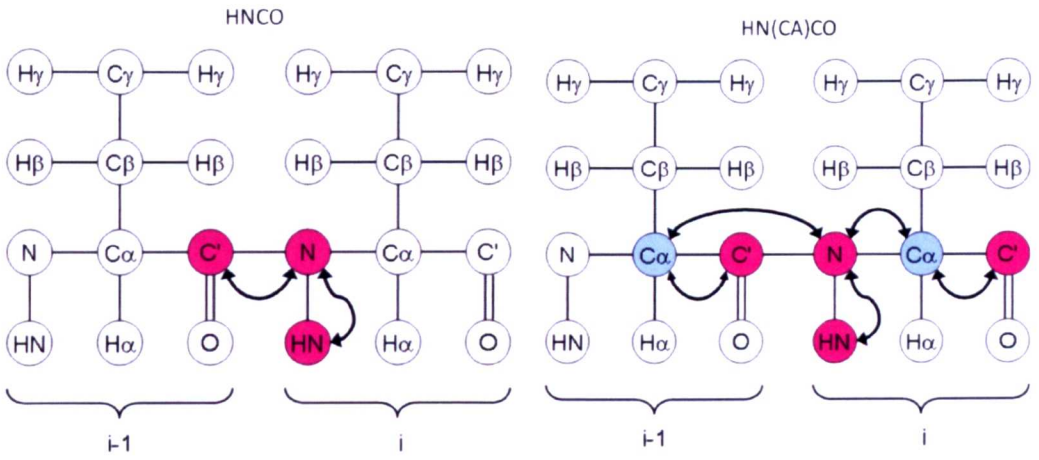


Figure 2.9 – The transfer of magnetization for the HN(CA)CO and HNCO triple resonance NMR experiments which were used in this thesis. The transfer of magnetization in the experiments used to identify the chemical shifts of C'[209].

The chemical shifts for the Cα and Cβ resonances are used for the sequential assignment of proteins. The range of chemical shifts for Cα and Cβ nuclei for different amino acids is quite narrow in the carbon dimension. However, a few residues are quite distinctive which can be used as starting points in a sequential assignment strategy (figure 2.9 A). Gly residues, as indicated earlier, are distinctive because they lack a Cβ resonance. Ser and Thr residues have Cβ resonances located above 60 ppm, which is much higher than other amino acids. Moreover, Ala residues have Cβ resonances located below 22 ppm. All other amino acids have Cβ resonances located between 25 and 45 ppm (approximately). A chart highlighting the typical chemical shift range for Cα and Cβ resonances is shown in figure 2.9 B.

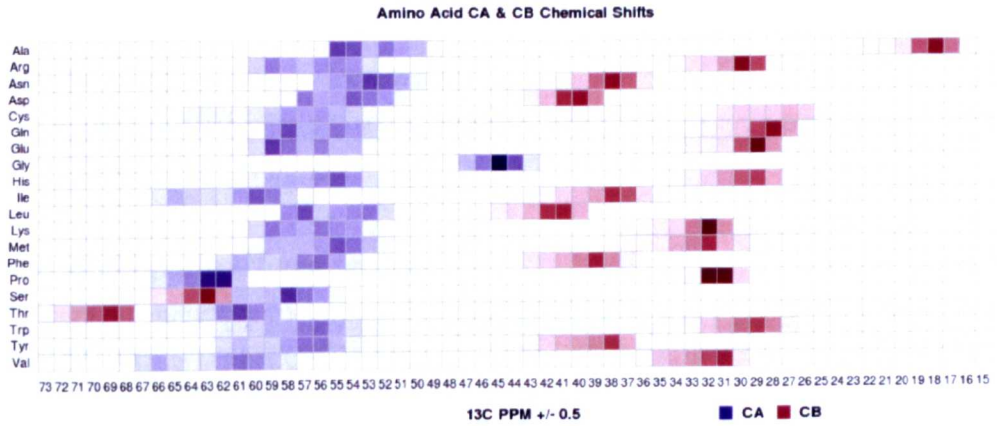


Figure 2.10 –the average chemical shifts for Cα and Cβ resonances for all the amino acids. The Cα and Cβ ranges are indicated in blue and red respectively[212].

3.0 Materials and methods

The proteins listed below were used in the studies detailed in this thesis. The gene segments for the full length p62 (residues 1-440), the p62 subfragment encoding residues 341-440, p62 UBA (residues 387-436) and hHR23A UBA2 (residues 319-363) were previously cloned into the pGEX-4T-1 plasmid by the group of Dr Robert Layfield (School of Biomedical Sciences, University of Nottingham). The MAP-LC3 protein was previously cloned into the pGEX-4T-3 plasmid also by the Layfield group. The p62 subfragments encoding residues 261-440 and 300-440 were personally cloned into the pGEX-4T-1 plasmid and confirmed by DNA sequencing. Proteins cloned into pGEX plamsids are cleaved from a GST tag leaving a short extension at the start of the sequence (see table 3.6). The gene for yeast ubiquitin was cloned into the pkk223-3 plasmid previously by Dr Geoff Platt (formerly of the Searle group). Linear diubiquitin was cloned into the pt7-7 plasmid by Paul Sheppard (formerly ENZO Life Sciences). Once cloned into the various plasmids the genes were transformed into suitable *Escherichia coli* (*E. coli*) cell lines for subsequent growth and purification protocols.

p62 (*Homo sapiens*)

MASLTVKAYLLGKEDAAREIRRFSFCCSPEPEAEAEAAAGPGPCERL
LSRVAALFPALRPGGFQAHYRDEDGDLVAFSSDEELTMAMSYVKDDI
FRIYIKEKKECRRDHRPPCAQEAPRNMVHPNVICDGCNGPVVGTRYK
CSVCPDYDLCSVCEGKGLHRGHTKLAFPSPFGHLSEGFSHSRWLRKV
KHGHFGWPGWEMGPPGNWSPRPPRAGEARPGPTAESASGPSEDPSVN
FLKNVGESVAAALSPLGIEVDIDVEHGGKRSRLTPVSPESSSTEEKS
SSQPSSCCSDPSKPGGNVEGATQSLAEQMRKIALESEGRPEEQMESD
NCSGGDDDWTHLSSKEVDPSTGELQSLQMPESEGPSSLDPSQEGPTG
LKEAALYPHLPPEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYD
IGAALDTIQYSKHPPPL*

Ubiquitin (*Saccharomyces cerevisiae*)

MQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAG
KQLEDGRTLSDYNIQKESTLHLVLRLRGG*

Linear diubiquitin (*Homo sapiens*)

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAG
KQLEDGRTLSDYNIQKESTLHLVLRLRGGMQIFVKTLTGKTITLEVE
PSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKES
TLHLVLRLRGG*

hHR23A UBA2 (*Homo sapien*)

QEKEAIERLKALGFPESLVIQAYFACEKNENLAANFLLSQNFDDE*

MAP-LC3 (*Homo sapien*)

MPSEKTFKQRRTFEQRVEDVRLIREQHPTKIPVIIERYKGEKQLPVL
DKTKFLVPDHVNMSELIKIIRRRLQLNANQAFFLLVNGHSMVSVSTP
ISEVYESEKDEDGFLYMVYASQETFGMKLSV

3.1 Materials

All chemicals were purchased from Sigma, Melford or Fisher chemical companies and were supplied at analytical grade

3.2 Molecular biology techniques

3.2.1 Sterilisation

A sterile environment was maintained during the growth of all proteins. All bench surfaces were decontaminated regularly with Trigene. Before use 100% ethanol was used to clean the surfaces further and all growth experiments were conducted in close proximity to a strong blue Bunsen flame. Disposable equipment (petri dishes, 20 ml universals, syringes, needles and filters) were supplied sterile. Pipette tips, eppendorfs, growth media and all buffers were sterilised by heating to 121 °C in an autoclave (Philip Harris, Status Autoclave) before use. Heat sensitive reagents were filtered using a 0.22 μm filter (sartorius stedim) before use in the purification protocol.

3.2.2 Luria Agar plates

Luria Agar powder (40 g/L) was dissolved in Milli-Q water and sterilised in an autoclave. Sterile solidified agar was then heated in short bursts until melted. The liquid was then cooled under a cold tap to approximately 40 °C for addition of sterile filtered antibiotics. Ampicillin at a final concentration of 270 μM was added to select for cells with ampicillin resistance. All plasmids used in these experiments had ampicillin resistance. Addition of tetracycline at a final concentration of 140 μM was required for the XL1-Blue strain of *E.coli* (XL1-blue are tetracycline resistant). The agar was poured into sterile Petri dishes, sealed and allowed to set at room temperature. Unused plates were stored at 4 °C until needed.

Solutions

Luria Agar (Sigma): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar, autoclaved at 121 °C for 15 minutes.

3.2.3 Overnight *E.coli* cultures

Individual isolated *E.coli* colonies (C41 (DE3) or XL1-blue) grown on Luria agar plates were used to inoculate 10 ml Luria Broth (LB) containing the appropriate antibiotic selections. The cultures were incubated at 37 °C with 180 rpm agitation for a minimum of 16 hours. A control containing the growth medium and the antibiotics alone was also prepared. The fully grown 10 ml culture was used to create glycerol stocks which stored several copies of the plasmid containing the gene of interest. Subsequent inoculations therefore involved scraping cells from the glycerol stock using a sterile pipette tip instead of obtaining the cells from the agar colony.

Solutions:

LB broth (Sigma): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, autoclaved at 121 °C for 15 minutes.

3.2.4 Glycerol stocks

Aseptic 30 % (w/v) glycerol and cells from an overnight culture were added to a sterile eppendorf at a 1 : 1 ratio and sealed under sterile conditions. The new stocks were stored at -80 °C until required for inoculation of overnight cultures.

3.2.5 Plasmid purification

High quality plasmid DNA was purified from XL1-blue *E.coli* cells containing the gene of interest. Overnight cell cultures containing tetracycline (selective for XL1- blues) and ampicillin (selective for the plasmid) were cultivated. The cells were harvested by centrifugation at 1000 rcf for 10 minutes and the resultant supernatant was discarded. The plasmid DNA was subsequently purified using a QIAprep® Miniprep kit (Qiagen) as per the manufacturer's instructions. DNA is lysed from the bacterial cells using an alkaline lysis methodology. Precipitation of cellular debris and genomic DNA occurs after the lysis reaction is neutralised. Separation of the plasmid DNA from the precipitate was achieved using centrifugation. The plasmid DNA is absorbed into a silica membrane and eluted

from the membrane in 50 µl of elution buffer (EB) and stored at -20 °C until required.

Solutions:

EB Buffer: 10 mM Tris.HCl (pH 8.5) (Qiagen)

3.2.6 Agarose gel electrophoresis

Agarose gels were used to separate DNA molecules according to their molecular weight. Mini-sub® cell GT apparatus (BioRad) was assembled as per the manufacturer's instructions. Agarose gels, 0.5-2% (w/v), were generated by dissolving agarose in 42 ml Tris Acetate EDTA (TAE) buffer. The solution was heated gently until boiling. Ethidium bromide, a DNA intercalator, was then added at a final concentration of 0.1 µl/ml to visualise the DNA. The heated solution was poured into the assembled apparatus, a comb inserted and the solution allowed to cool. Once fully set the comb was removed and the gel submerged in TAE buffer. DNA samples were treated with 6 x loading dye (Promega), 1 µl dye to 5 µl DNA, and loaded onto the gel. A 70 Volts (V) potential was applied until the bands had separated sufficiently (20-60 min). Bands were visualised under a GBOX UV transluminator (Syngene) and their molecular weights determined in comparison to a 2-log ladder (NEB) used as a marker.

Solutions

TAE buffer: 40 mM Tris.HCl, 20 mM acetic acid, 1 mM EDTA (pH 5.0)

2-log loading ladder: 0.1-10kb containing 19 differently sized fragments used as molecular weight standards

3.2.7 DNA concentration

DNA concentration was measured using a Nanodrop ND-100 spectrophotometer (NanoDrop). The nanodrop measures the ultraviolet (UV) absorbance at 260 nm for DNA samples. The absorbance at 260 nm was then entered into the Beer-Lambert equation to calculate concentration.

Beer-Lambert law $A = \varepsilon c l$

Where A is absorbance (A_{260}), c is concentration (M), ε is the molar extinction coefficient and l is the path length (cm). The molar extinction coefficient of double stranded DNA is taken as $\varepsilon_{260} = 0.02$ $(ng/\mu l)^{-1}$ cm^{-1}[213].

3.2.8 Cloning the p62 subfragments into the PGEX-4T-1 plasmid

3.2.8.1 PCR gene amplification

The gene of interest was amplified by PCR in preparation for its insertion into the pGEX-4T-1 plasmid. Oligonucleotide primers were synthesised by the School of Biomedical Sciences, University of Nottingham. Primers were designed according to standard guidelines. The p62 subfragment encoding residues 261-440 was cloned between the *EcoRI* and *XhoI* restriction sites; whereas, the subfragment encoding residues 300-440 was cloned between the *BamHI* and *XhoI* restriction sites. The details of the forward and reverse primers are listed in the appendix.

Plasmid template (50 ng μl^{-1}, 1 µl), 10 x reaction buffer (5 µl, Stratagene), excess forward and reverse primers (125 ng μl^{-1}, 1 µl each primer) and dNTP mix (10 mM each nucleotide, 1 µl, Stratagene) were added to a sterile PCR tube and placed on ice. *Pfu Turbo*® DNA polymerase (2.5 U μl^{-1}, 1 µl, Stratagene) was added to the reaction last to make the total volume 50 µl. The reaction was transferred to a thermocycler (Techgene) according to the following program. The PCR reaction utilised the parental DNA denauration, primer annealing and primer extension stages. The number of cycles of the PCR was increased for gene amplification when compared to site directed mutagenesis to amplify sufficient DNA for cloning experiments.

1 cycle	95 °C	for	30sec
30 cycles	95 °C	for	30sec
	60 °C	for	30sec
	72 °C	for	30sec
1 cycle	4 °C	for	∞

Solutions:

10 X Reaction Buffer: 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 0.1% Triton® X-100, 0.1mg/ml^{-1} nuclease free bovine serum albumin (BSA).

3.2.8.2 Vector restriction digests

The plasmid DNA was prepared using a miniprep kit as outlined in section 3.2.5. The pGEX-4T-1 was digested with the appropriate restriction endonucleases (Promega) as reactions summarised in table 3.1. Single digests required the use of the buffer supplied with that enzyme. Double digests buffers were calculated using the promega website (see table 3.2) to maximise the enzymatic activity of both endonucleases. Uncut DNA and single cut controls were also prepared to verify that all enzymes were cutting. The reactions were completed in a total volume of 40 µl and incubated at 37 °C for 4 hours. The restriction endonucleases were denatured at 65 °C for 15 minutes upon reaction completion. To prevent self ligation of the newly generated sticky ends, Antarctic phosphatase (Biolabs) was added to each reaction at a concentration of 5 units/µl (see table 3.3). The total reaction volume was increased to 50 µl with the addition of the phosphatase, phosphatase buffer and water. The reaction returned to 37 °C for 10 minutes. The antarctic phosphatase was also heat inactivated as described for the restriction endonucleases. The digested products were separated from the denatured enzymes on a 1% (w/v) agarose gel as outlined in section 3.2.6, visualised using a UV light source and excised from the gel for further purification.

Solutions

Buffer E: 60 mM Tris.HCl (pH 7.5), 1 M NaCl, 60 mM $MgCl_2$, 10 mM DTT

Buffer H: 900 mM Tris.HCl (pH 7.5), 500 mM NaCl, 100 mM $MgCl_2$

Buffer D: 60 mM Tris.HCl (pH 7.9), 1.5 M NaCl, 60 mM $MgCl_2$, 10 mM DTT

Buffer C: 100 mM Tris.HCl (pH 7.9), 500 mM NaCl, 100 mM $MgCl_2$, 10 mM DTT

Antarctic phoshatase buffer: 50 mM Bis Tris-Propane, (pH 6.0), 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$

Table 3.1 – The vector restriction digest conditions used in this study. Each reaction has a total volume of 40 µl.

Reaction composition (µl)	**Uncut plasmid control**	**Single cut BamHI**	**Single cut EcoRI**	**Single cut XhoI**	**Double digest p62 300 - 440 (BamHI + XhoI)**	**Double digest p62 261 – 440 (EcoRI + XhoI)**
Milli-Q water (µl)	28.5	26.5	26.5	24.5	22.5	22.5
10 x buffer (µl)	4	4	4	4	4	4
Plasmid DNA (120ng/µl) (µl)	6.5	6.5	6.5	6.5	6.5	6.5
50 x BSA (µl) (µl)	1	1	1	1	1	1
BamHI (10 units/µl) (µl)	N/A	2	N/A	N/A	2	N/A
EcoRI (12 units/µl) (µl)	N/A	N/A	2	N/A	N/A	2
XhoI (10 units/µl) (µl)	N/A	N/A	N/A	4	4	4

Table 3.2 – The 10 X buffer used in vector restriction digests. The buffer was chosen to maximise the activity of all enzymes contained within the reaction.

Reaction	**10 x buffer used**
Uncut plasmid control	Buffer E
Single cut BamHI	Buffer E
Single cut EcoRI	Buffer H
Single cut XhoI	Buffer D
Double digest BamHI + XhoI	Buffer C
Double digest EcoRI + XhoI	Buffer H

Table 3.3 – Reagents added to vector restriction digest to prevent self ligation

Reagent added to 40µl restriction digest reaction
5µl 10 x antarctic phosphatase buffer
4µl Milli-Q water
1µl antarctic phoshpatase (5 units/µl)

3.2.8.3 Insert restriction digests

The PCR amplified gene inserts were gel extracted (see section 3.2.8.4) and the concentrations measured using a NanoDrop ND-100 spectrophotometer (NanoDrop). The inserts were subsequently digested under identical conditions to those used in the vector restriction digests. Conditions include the Molarity of DNA being digested, units of restriction endonuclease and time of digestion. The reactions were set up as outlined in table 3.4 and incubated at 37 °C for 8 hours.

Only double digests were completed because the restriction sites are located at the ends of the amplified DNA insert. Digestion therefore results in a very small size change in the DNA fragment which is unable to be detected by agarose gel. Buffers were chosen for the double digestions as outlined in table 3.2 (section 3.2.8.2).

Solutions

Buffer H: 900 mM Tris.HCl (pH 7.5), 500 mM NaCl, 100 mM $MgCl_2$

Buffer C: 100 mM Tris.HCl (pH 7.9), 500 mM NaCl, 100 mM $MgCl_2$, 10 mM DTT

Table 3.4 – Insert restriction digests. The difference in the volume of insert DNA is based on the concentration of the insert as determined by the nanodrop.

Reaction composition (µl)	**Double digest p62 300 - 440 (BamHI + XhoI)**	**Double digest p62 261 – 440 (EcoRI + XhoI)**
Milli-Q water (µl)	18	34
10 x buffer (µl)	5	5
Insert DNA (12ng/µ for 300-440) (42ng/µl for 261-440) (µl)	20	4
50 x BSA (µl)	1	1
BamHI (10 units/µl) (µl)	2	N/A
EcoRI (12 units/µl) (µl)	N/A	2
XhoI (10 units/µl) (µl)	4	4

3.2.8.4 Gel extraction purification

DNA purification of fragments 70 base pairs (bp) to 10 kilo bases (kb) from enzymatic reactions were performed using a QuikChange ® gel extraction kit (Qiagen) as per the manufacturer's instructions. The step was necessary after enzymatic reactions (PCR gene amplification and restriction digests for both vector and insert DNA) in order to obtain high quality pure plasmid DNA by removing unwanted primers, enzymes and other impurities. The plasmid DNA is isolated using agarose gel electrophoresis. The band of interest is excised from the gel and is subsequently solubilised using a buffer containing the chaotrophic NaCl guanidiunium thiocyanate. In this environment the plasmid DNA preferentially binds to the QIAquick silica membrane. The DNA is washed and eluted in 50 µl

using a buffer with low NaCl under basic conditions. The eluted pure DNA is stored at -20 °C until required.

Solutions:

EB Buffer:10 mM Tris.HCl (pH 8.5) (Qiagen)

3.2.8.5 Ligations

The digested vector and digested insert were ligated together using a T4 DNA ligase (Promega). A 4:1 insert:vector ratio was used to ensure the ligation was successful. The total volume of insert and vector did not exceed 11 µl, with the total volume of each reaction being 20 µl (table 3.5). A control reaction containing only vector was also set up. The ligations were completed at 16 °C for 16 hours. The T4 DNA ligase was denatured at 65 °C after reaction completion.

Table 3.5 – Ligation reaction compositions. The same control was used for both reactions as they both contained the same content of vector DNA.

	p62 300-440	**p62 261-440**	**Control**
Vector DNA (7ng/µl for 300-440) (9ng/µl for 261-440) (µl)	4	4	4
Insert DNA (4ng/µl for 300-440)(6ng/µl for 261-440) (µl)	6	7	N/A
Milli-Q water (µl)	7	6	13
T4 ligase buffer (µl)	2	2	2
T4 ligase (3 units/µl) (µl)	1	1	1

Solutions

T4 DNA ligase buffer: 300 mM Tris.HCl (pH 7.8), 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP

3.2.9 Site directed mutagenesis

3.2.9.1 Polymerase chain reaction (PCR)

All p62 UBA mutants and the C331S p62 300-440 mutant (outlined in section 1.6) were generated using a QuikChange® Site directed mutagenesis kit (Stratagene). Oligonucleotide primers containing the desired mutations were designed according to Quikchange guidelines and synthesised by the School of Biomedical Sciences, University of Nottingham. The details of the forward and reverse primers are listed in the appendix. A 50 ml reaction was prepared in a thin walled 0.2 ml PCR tube on ice. Plasmid template (50 ng μl^{-1}, 1 µl), 10 x reaction

buffer (5 µl, Stratagene), excess forward and reverse primers (125 ng μl^{-1}, 1 µl of each primer) and dNTP mix (10 mM each nucleotide, 1 µl, Stratagene) were added to the sterile PCR tube. *Pfu Turbo*® DNA polymerase (2.5 U, 1 µl, Stratagene) was added last and the reaction was transferred to a thermocycler (Techgene) according to the program detailed below. PCR consists of 3 main stages, parental DNA denaturation, primer annealing and primer extension.

1 cycle	95 °C	for	30sec
n cycles	95 °C	for	30sec
	55 °C	for	60sec
	68 °C	for	1 min/kb plasmid length
1 cycle	4 °C	for	∞

(point mutations n = 12, single amino acid changes n = 16, multiple amino acid changes n = 18).

Solutions:

10 X Reaction Buffer: 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 0.1% (v/v) Triton® X-100, 0.1mg/ml-1 nuclease free bovine serum albumin (BSA).

3.2.9.2 Dpn1 Digestion

All mutagenesis PCR reactions were treated with restriction enzyme *Dpn1*. *Dpn1* (0.2 U µl-1) was added to the PCR reaction tube and heated to 37 °C for 1 hour. Parental DNA is digested by *Dpn1*, which is selective for methylated or hemimethylated DNA. The remaining undigested DNA should therefore contain the desired mutation.

3.2.10 Transformation into XL-1 blue *E.coli*

3.2.10.1 Preparation of chemical supercompetent cells

Glycerol stocks of XL1 blue *E.coli* were used to inoculate 10 ml LB without antibiotic selection and left to grow at 37 °C for 16 hours with 180 rpm agitation. These cells were used to inoculate another 10 ml LB at 1 : 50 dilution and grown until the OD_{595} was 0.3. Cells were spun at 1000 rcf in a centrifuge for 10 minutes, resuspended in 4 ml 50 mM $CaCl_2$ and left on ice for 3 hours. The cells

were centrifuged again (as above), resuspended in 800 µl 50 mM $CaCl_2$ and left for 30 minutes to make them chemically competent.

3.2.10.2 Heat shock transformation of DNA into host cells

Chemically competent XL1-blue aliquots (100 µl) were treated with 1 µl or 3 µl DNA from site directed mutagenesis reactions or ligation reactions respectively. The aliqouts were then chilled on ice for 30 minutes. The cells were subsequently heat shocked at 42 °C for 30 seconds to uptake the plasmid DNA and returned to ice for a further 2 minutes. A small volume of LB (400 µl) was added to the aliquots and they were allowed to grow at 37 °C for 1 hour. The cells were spread onto Agar plates containing ampicillin (270 µM) and tetracycline (140 µM) and incubated at 37 °C for 16 hours or until colonies had grown.

The *E.coli* strain XL1-blue were chosen because they can repair nicked DNA. XL1-blue are endonuclease A (*endA*) and recombination A (*recA*) deficient. *EndA* deficiency enables high quality DNA to be produced from miniprep kits; whereas, *recA* deficiency increases the stability of the inserted DNA within the plasmid.

3.2.11 DNA sequencing

All newly cloned plasmids and PCR mutants were verified by DNA sequencing. The DNA was purified using a miniprep kit (section 3.2.5) and taken to the DNA sequencing facility (School of Biomedical sciences, University of Nottingham) for confirmation. A 5µl DNA sample (40 – 100 ng/µl) was sufficient for the sequencing reaction and the forward and reverse primers (both > 10 pmol/µl) were supplied by the DNA sequencing facility. The DNA was sequenced using the standard Sanger method[214]. A chromatogram was produced and visualized using Chromas 2.21(Technelysium).

3.2.12 Transformation into C41 (DE3) *E.coli*

Overnight cultures containing no antibiotic selection were inoculated with C41 (DE3) glycerol stocks. Chemically competent cells were prepared as outlined in 2.2.10.1. Transformations were completed as outlined in 2.2.10.2 but with agar plates selective for ampicillin (270 µM) only.

3.3 Protein overexpression

3.3.1 Overexpression plasmids

The details of the plasmids and proteins used in this study are summarised in table 3.6. The restriction sites the genes are located between and the amino acid extension left from thrombin cleavage, where applicable, are also indicated. The plasmid maps of all plasmids used throughout this study can be found in the appendix.

The pGEX expression system (GE healthcare) is a highly efficient system which is used to express glutathione S-transferase (GST) fusion proteins. The gene which codes for the GST protein from *Schistosoma japonicum* (bloodworm) had been previously cloned into the PGEX plasmids by the supplier. The gene encoding the protein of interest was cloned downstream of the GST gene in the multiple cloning site. A *tac* promoter, a functional hybrid derived from the *trp* and *lac*UV5 promoters[215], is located upstream of the gene which encodes the GST tag and the multiple cloning site.

Overexpression of the desired protein is regulated by the *tac* promoter which is controlled by T7 RNA polymerase. Under normal conditions gene expression is driven by the addition of a lactose metabolite called allolactase, which binds to the lac repressor. When isopropyl-thio-β-D-galatoside (IPTG), a non hydrolysable β galactosidase analogue, is added it binds to the lac repressor protein in both the bacterial host DNA and the pGEX plasmid DNA. When IPTG binds to the lac repressor in the bacterial DNA the *lac*UV5 promoter drives the expression of T7 RNA polymerase from the bacterial genome. In the plasmid DNA IPTG also binds to the lac repressor protein allowing the T7 RNA polymerase to bind to the *tac* promoter to initiate transcription of the desired gene. The T7 RNA polymerase is not produced by the bacterial cells therefore the bacterial genome needs to be modified to encode the gene for the T7 RNA polymerase.

The pGEX-4T-1 and pGEX-4T-3 plasmids, as used in this study, express GST fusion proteins. The GST protein is a very soluble protein capable of enhancing the solubility of the fused protein. Separation of the fusion protein is achieved using the proteolytic enzyme thrombin. The thrombin recognition sequence, Leu-

Val-Pro-Arg-Gly-Ser, is located between the GST and the cloned protein of interest. Thrombin cleaves between the Arg and Gly residues leaving a minimum of a Gly-Ser extension after cleavage. The extension length is dependent on the restriction sites the gene of interest was cloned into within the multiple cloning site. The details of the restriction sites and the thrombin cleavage extension used for each protein are indicated in table 6.

The pKK233-3 also possesses a *tac* promoter upstream of the gene of interest; however, the pt7-7 plasmid contains a T7 promoter. The T7 promoter also permits high efficiency transcription[216]. The desired gene is transcribed from the T7 promoter by T7 RNA polymerase in much the same way as the *tac* promoter does. The main difference between these plasmids and the PGEX expression system is that these plasmids lack a GST affinity tag. Ubiquitin is a highly studied protein and is known to be soluble at a variety of pH's. An affinity tag to aid protein solubility was therefore not required.

Table 3.6 – A table summarising the plasmid details of all the proteins used in this study

Protein	Plasmid	Restriction sites	Affinity tag	Extension
P62 341-440	pGEX-4T-1	EcoRI-XhoI	GST	GSPEF
P62 300-440	pGEX-4T-1	BamHI-XhoI	GST	GS
P62 261-440	pGEX-4T-1	EcoRI-XhoI	GST	GSPEF
P62 UBA	pGEX-4T-1	BamHI-XhoI	GST	GS
MAP-LC3	pGEX-4T-3	EcoRI-SalI	GST	GSPNS
hHR23A UBA2	pGEX-4T-1	BamHI-XhoI	GST	GS
Ubiquitin	pKK233-3	EcoR1-HindIII	None	None
Linear diubiquitin	pt7-7	NdeI-EcoRI	None	None

3.3.2 Protein overexpression using C41 (DE3) *E.coli*

The *E. coli* strain C41 (DE3) was modified from BL21 (DE3) strain in order to express proteins which are toxic to the BL21 (DE3) cells[217, 218]. C41 (DE3) *E. coli* have been shown to be effective in expressing recombinant proteins from all classes of organisms. This strain has at least one uncharacterised mutation, which prevents the cell death associated with the expression of toxic recombinant proteins. The λ prophage DE3, which encodes for the T7 RNA polymerase after a *lac*UV5 promoter, had been previously inserted into the BL21 bacterial DNA. The *lac*UV5 promoter is recognised by the *E. coli* RNA polymerase and is induced by the presence of IPTG. Also like the parental BL21 (DE3) strain, C41 (DE3) is

deficient in *Ion* and *ompT* proteases which allows for better isolation of intact recombinant proteins. C41 (DE3) cells possess no intrinsic antibiotic resistance.

3.3.3 Growth media

3.3.3.1 LB media

Luria broth (LB) medium was used for the expression of unlabelled proteins, except ubiquitin. 25 g LB powder was dissolved in 1 litre of Milli-Q water and transferred to a 2 litre baffled flask. The media was subsequently autoclaved to sterilise.

Solutions:

LB broth (Sigma): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, autoclaved at 121 °C for 15 minutes.

3.3.3.2 M9 minimal media

M9 minimal media was used to overexpress unlabelled ubiquitin and all ^{15}N and $^{13}C/^{15}N$ labelled proteins. M9 minimal media was prepared in a 2 litre baffled flask by dissolving M9 NaCls in 1 litre of Milli-Q water and autoclaved to sterilise. The media was supplemented with 20 ml of sterile filtered (0.22 μm sartorius stedim) M9 nutrient solution and biotin and thiamine to a final concentration of 10 mg/l after sterilisation.

Solutions:

M9 NaCls: Calcium chloride dehydrate (15 mg/l), Di-Sodium orthophosphate (6 g/l), Magnesium sulphate (0.5 g/l), Potassium dihydrogen orthophosphate (3 g/l), NaCl (1 g/l).

M9 nutrient solution: 1 g Ammonium Chloride, 4 g glucose in 20 ml Milli-Q water.

3.3.3.3 Isotopic labelling

For NMR experiments, proteins were isotopically labelled with either ^{15}N or a combination of ^{15}N and ^{13}C[219]. Thus ^{15}N-enriched ammonium chloride and ^{13}C-enriched glucose are required to replace the naturally abundant ^{14}N-ammonium chloride and ^{12}C-glucose components of the M9 nutrient solution. The ammonium chloride and glucose were the sole source of nitrogen and carbon respectively

meaning that proteins expressed from this media will possess the appropriate isotopic labelling. The media was prepared and autoclaved as normal. The isotopically labelled M9 nutrient solution was added along with ampicillin, biotin and thiamine as previously outlined. For $^{15}N/^{13}C$ proteins the amount of ^{13}C-glucose was halved.

Solutions:

^{15}N nutrient solution: 1 g ^{15}N-ammonium Chloride, 4 g glucose in 20 ml Milli-Q water.

$^{15}N/^{13}C$ nutrient solution: 1 g ^{15}N-ammonium Chloride, 2 g ^{13}C-glucose in 20 ml Milli-Q water.

3.3.4 Small scale protein overexpression

An overnight culture was inoculated with C41 (DE3) *E.coli* containing the desired protein and was grown at 37 °C with 180 rpm agitation for 16 hours. The overnight culture was used to inoculate 10 ml LB containing ampicillin (270 μM) at a 1 in 50 dilution and grown at 37 °C with 180 rpm agitation until the OD_{595} was 0.7 (1-3 hours). Protein expression was induced by the addition of IPTG at a final concentration of 1 mM. The culture transferred to 30°C for protein expression. A 1 ml sample was taken prior to IPTG addition. Further 1 ml samples were taken at 2, 3, 4, 5, 6 and 16 hours after protein induction and the OD_{595} checked. The samples were centrifuged at 13,000 rcf for 10 minutes and the supernatant discarded. The resultant cell pellet was resuspended in 300 μl lysis buffer A for proteins expressed using the pGEX expression system and 5 x AE for monoubiquitin and linear diubiquitin. The 1 ml samples were sonicated in a soniprep 150 sonicator (MSE) for 10 seconds (x2) at 10 microns. The sonicated samples were centrifuged once more. Samples (20 μl) were taken before (total cell lysate) and after (soluble fraction) centrifugation and visualised on an SDS PAGE gel to observe protein expression

Solutions

Lysis buffer A: 10mM Tris.HCl, 150mM NaCl (pH 7.4)

5 x AE buffer: 250 mM acetic acid, 10 mM EDTA, pH 4.0

3.3.5 Large scale overexpression

Once successful overexpression had been observed on a small scale, large scale growths ensued. Overnight cultures were made in LB media. Glycerol stocks containing the desired plasmid in C41 (DE3) *E.coli* were used to inoculate the overnight cultures. Large baffled flasks containing either 1 litre of sterilised LB or M9 minimal media were inoculated with overnight cultures. Each litre of autoclaved LB was inoculated with 20 ml overnight culture whereas each litre of autoclaved M9 minimal media was inoculated with 45 ml (3 x 15 ml) overnight cultures. Cells used to inoculate LB were added directly to the LB media. Cells used to inoculate M9 minimal media were centrifuged at 1000 g for 10 minutes and the supernatant discarded. The cell pellets were resuspended in M9 minimal media which had been enriched with the M9 nutrients, biotin and thiamine. The cells were returned to 37 °C and shaken at 180 rpm for 30 minutes before adding directly to M9 minimal media. Ampicillin at a final concentration of 270 μM was added to the media at inoculation as all proteins expressed in this study possessed ampicillin resistance.

The inoculated growths were then left to grow at 37 °C with 180 rpm agitation until the OD_{595} was 0.7. Prior to protein induction the incubator was cooled to 30°C. The cells were induced with IPTG at a final concentration of 1 mM and returned to the incubator to induce at 30°C. A further 270 μM ampicillin was added at induction to replace the ampicillin metabolised by the *E.coli*. Proteins expressed in LB were induced for 4-10 hours (table 7). Proteins expressed in M9 minimal media were induced for a minimum of 16 hours (table 3.7). When the OD_{595} approached 1.8 the cells were harvested by centrifugation (Sorvall) at 3000 rcf for 10 minutes. The resultant cell pellet was scraped into a Falcon tube. The remaining cells were resuspended in lysis buffer A for proteins expressed using the pGEX expression system and 5 x AE buffer for monoubiquitin and linear diubiquitin and added to the Falcon tube containing the cell pellet. The cells were centrifuged at 5000 rcf for 15 minutes, the supernatant discarded and the pellet stored at -20 °C until needed. 1 ml gel samples were taken before induction and harvesting. The gel samples were spun, resuspended and sonicated as described in the small scale overexpressions.

Table 3.7 – The growth conditions for all the proteins used in this study.

Protein	Growth temperature (°C)	Induction temperature (°C)	Induction time in LB (hours)	Induction time in M9 (hours)
P62 341-440	37	30	8	16
P62 300-440	37	30	8	16
P62 261-440	37	30	8	16
P62 UBA	37	30	8	16
MAP-LC3	37	30	10	N/A
hHR23A UBA2	37	30	4	16
Ubiquitin	37	30	N/A	16
Linear diubiquitin	37	30	8	16

Solutions

Lysis buffer A: 10mM Tris.HCl, 150mM NaCl (pH 7.4)

5 x AE buffer: 250 mM acetic acid, 10 mM EDTA, pH 4.0

3.3.6 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gel apparatus (Biorad) were assembled as per the manufacturer's instructions. The gel contained a 20% acrylamide resolving layer beneath 4% acrylamide stacking layer. The resolving layer consisted of a 20% acrylamide stock containing ammonium persulfate (APS) (0.5% v/v) and N,N,N,N-tetramethylethylethylenediamine, (TEMED) (1% v/v). Additon of APS and TEMED catalysed the formation of a solid gel. The stacker, consisting of a 4% acrylamide stock solution, contained identical concentrations of APS and TEMED to those described for the resolving layer. The stacking layer was prepared once the resolving layer had polymerised. A comb was inserted quickly to the setting stacking layer to form the loading wells. Protein samples, 20 µl, were mixed with 10 µl 2 x protein loading dye, containing 20% DL-dithiothreitol (DTT) for sample denaturation. The samples were heated at 90 °C for 2 minutes. The gel assembly containing the set gel was transferred to the electrophoresis pack, the area between the two gels filled with 1 x SDS and the combs removed. A low molecular weight marker (Sigma) or Novex prestained protein standard marker (Invitrogen) was loaded as a size standard and 8 µl samples were subsequently loaded. The gel was run at 200 mV, immersed in 1 x SDS buffer, for 1 hour. Proteins were visualised by Coomassie Blue stain and destained with 25% acetic

acid and 10% methanol. The protein bands were visualised and photographed using a transilluminating light box.

Solutions:

TE buffer: 10 mM Tris-HCl (pH 7.4), 1 mM EDTA.

20 % Acrylamide solution: 20% Protogel® (20 % (w/v) acrlamide, 0.4 % (w/v) bis-acrylamide (National Diagnostics), 375mM Tris-HCl (pH 8.8), 0.1 % SDS.

Stacking solution: 2 % Protogel® (National Diagnostics), 125 mM Tris-HCl (pH 6.8), 0.1 % (w/v) SDS.

2 x protein loading buffer: 100 mM Tris-HCl (pH 6.8), 4 % (w/v) SDS, 20 % (w/v) glycerol, 0.2 % (w/v) bromophenol blue, 200 mM dithiolthreitol (DTT).

Low molecular weight marker (Sigma): albumin (66 kDa), ovalbumin (45 kDa), glyceraldehydes-3-phosphate dehydrogenase (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (24 kDa), trypsin inhibitor (20kDa), α-lactalbumin (14.2 kDa) and aprotinin (6.5 kDa).

Novex prestained protein standard marker (Invitrogen): 12 pre-stained protein bands in the molecular weight range of 3.5 kDa to 260 kDa.

SDS running buffer: 25 mM Tris base, 250 mM glycine (pH 8.3), 0.1 % (w/v) SDS.

Staining Solution: 0.25 % (w/v) Brilliant Blue R-250 (Fisher), 45% (v/v) methanol, 10% (v/v) glacial acetic acid.

Destaining Solution: 25 % (v/v) methanol, 10% (v/v) glacial acetic

3.4 Protein Purification

Different columns attached to an ÄKTA prime (Amersham Pharmacia) FPLC system were utilised to purify the different proteins and their mutants. All proteins cloned into the pGEX-4T-1 and pGEX-4T-3 plasmids contain an N-terminal thrombin-cleavable GST tag. GST tagged proteins were subjected to a gravity flow talon column containing a glutathione Sepharose resin. The GST fusion proteins bind to the glutathione beads and are either cleaved from the beads using thrombin or eluted as a GST fusion protein. To elute from the column reduced

glutathione was applied to out compete the GST tag for binding to the beads. Both thrombin cleaved proteins and GST fusion were subjected to two further columns which separate proteins according to their size and shape. First a gel filtration column was utilised to remove any undesired products from the thrombin cleavage. Finally proteins were applied to a desalting column in order to remove any excess NaCl which could be detrimental to any subsequent NMR experiments. Nearly all proteins were lyophilised after the gel filtration and desalt columns to concentrate the protein. In the case of the MAP-LC3, the protein was concentrated using a spin concentrator between gel filtration and desalting steps.

Monoubiquitin (*Saccharomyces cerevisiae*) was previously cloned into the pKK223-3 plasmid whereas linear diubiquitin *(Homo sapien)*, was a kind gift from Paul Sheppard pre-cloned into the pt7-7 plasmid. Neither of these plasmids possesses a purification tag. Untagged monoubiquitin and linear diubiquitin were also purified using gel filtration and desalt columns but prior to gel filtration, a cation exchange column, which purifies protein based on charge, was used. Both monoubiquititn and linear diubiquitin are positively charged at pH 4, a pH where many proteins are insoluble. This fact makes these proteins relatively easy to isolate using a cation exchange column at this pH. Both proteins were lyophilised after the gel filtration and desalt stages of the purification protocol. Monoubiquitin was lyophilised after the cation exchange column; whereas, linear diubquitin was applied directly to the gel filtration column as 10 ml fractions.

3.4.1 Protein sonication and clarification

Frozen *E. coli* C41 (DE3) cell pellets containing the protein of interest were removed from the -80°C freezer and allowed to thaw gently. 100 µl aliquots of DNase I (Sigma) and protease inhibitor cocktail containing EDTA (Sigma) were added to the thawed pellet and the cells resuspended in the appropriate buffer. GST tagged proteins were resuspended in 6ml lysis buffer B containing DTT, EDTA and Triton X. Monoubiquitin and linear diubiquitin were resuspended in 20 ml per litre 5 x AE buffer. The cells were lysed by sonication at10 microns in an ice bath (30 sec bursts, 5 min total on time). Cellular debris was separated from cell lysate by two rounds of centrifugation at 4°C (40,000 rcf for 30 mins) using Avanti J-25 centrifuge with JL-20 rotor (Beckman). For monoubiquitin and linear diubiquitin a further 10 ml per litre 5 x AE buffer was added to the supernatant

between spins. Addition of 5 x AE buffer caused precipitation of impurities aiding the overall purification procedure as well as dragging the protein to pH 4 where the protein is positively charged.

Solutions

DNAse 1 Stock (Sigma):10 mg/ml (50, 000 U/ml) 1M $MgCl_2$, 0.1M $MnCl_2$.

Protease inhibitor cocktail (Sigma): AEBSF (2 mM), Aprotinin (0.3 μM), Bestatin (130 μM), EDTA (1 mM), E-64 (14 μM), Leupeptin (1 μM)

Lysis buffer B: 10 mM Tris.HCl, 150 mM NaCl (pH 7.5), 4 mM DTT, 15 mM EDTA, 0.1% (v/v) Triton X-100.

5 x AE buffer: 250 mM acetic acid, 10 mM EDTA, pH 4.0

3.4.2 GST Gravity Flow Talon Column

The supernatant of sonicated GST tagged proteins was applied to a column containing glutathione sepharose 4B beads (GE Healthcare). The sepharose beads were pre-equilibrated with 3 column volumes of lysis buffer A. The protein was tumbled slowly at 4 °C for 2-4 hours depending on protein stability. The tumbling process allowed GST tagged proteins to bind to the beads. Non-GST tagged proteins were collected in the flow through and washed out with 3 x column volumes (3 x 10ml) lysis buffer A washes. For GST cleaved purifications, proteins were subjected to an on-column cleavage by the protease thrombin. The column was initially washed with 3 column volumes of cleavage buffer to re-equilibrate the beads. The lid was then placed on the tip of the gravity flow column. Thrombin (5 units) was mixed with 990 μl thrombin cleavage buffer and applied to the column and tumbled at 4 °C for 16 hours. For GST fusion protein purifications 5 ml elution buffer was applied to the column after the lysis buffer A washes. The column was tumbled at 4 °C for a further 30 minutes. The glutathione contained in the elution buffer competes with the GST tagged protein for binding sites on the beads. A further 4 x 1 ml elution buffer is added to the column to remove any proteins still attached. The eluted protein was collected in a universal tube.

Solutions

Lysis buffer A: 10 mM Tris.HCl, 150 mM NaCl (pH 7.5)

Cleavage buffer: 20 mM Tris.HCl, 150 mM NaCl, 2.5 mM $CaCl_2$ (pH 8.4)

Elution buffer: 10 mM reduced glutathione, 50 mM Tris.HCl (pH 8.0)

3.4.3 Ion exchange chromatography

The supernatant from sonicated monoubiquitin and linear diubiquitin pellets was diluted to 150 ml with degassed Milli-Q water and filtered with a 0.45 μm filter (Minisart®, Sartorius) to remove any aggregates. The 150 ml solution was loaded onto a cation exchange column, pre-equilibrated with cation binding buffer, using a 50 ml superloop™ (Amersham Biosciences). Bound protein was eluted by increasing the NaCl concentration incrementally. Cation elution buffer was applied using a linear gradient. For monoubiquitin a sharp gradient of 0 -100% over 100 ml was applied. Linear diubiquitin utilised a more shallow gradient of 0-60 % over 500 ml. Both cation binding and elution buffers had been filtered with a 0.45 μm cellulose nitrate membrane (Whatman) and degassed before use. Fluorophores within the protein allowed the elutent to be visualised by absorbance at 280 nm. The protein was collected as 10 ml fractions. Fractions corresponding to the absorbance peak were confirmed by SDS-PAGE and combined.

Solutions:

Cation binding buffer: 50 mM acetic acid, 2 mM EDTA, pH 4.0, filtered and degassed

Cation elution buffer: 50 mM acetic acid, 2 mM EDTA, 2 M NaCl, pH 4.0, filtered and degassed

3.4.4 Gel filtration chromatography

Gel filtration is a technique which relies on separating proteins according to their size and shape. Larger proteins are eluted first whereas smaller, more compact proteins are able to weave between the pores of the column. A HiLoad™ Superdex™ 200 (Amersham biosciences) gel filtration column was used for all GST fusion proteins. A HiLoad™ Superdex™ 75 (Amersham biosciences) gel filtration column was used for monouniquitin, linear diubiquitin, C331S p62 300-

440 and MAP-LC3. A HiLoad™ Sephacryl™ S-100 (Amersham biosciences) gel filtration column was used for all thrombin cleaved p62 UBA and hHR23A UBA2 proteins. All gel filtration columns were pre-equilibrated with 3 x column volume (360 ml) potassium phosphate buffer before use. The phosphate buffer was filtered with a 0.45 μm cellulose nitrate membrane (Whatman) and degassed before column equilibration. Lyophilised monoubiquitin from the ion exchange column was resuspended in 2 ml Milli-Q water and loaded onto the gel filtration column. Linear diubiquitin was eluted in 10 ml fractions from the ion exchange column and loaded directly onto the gel filtration column using a 10 ml superloop (Amersham biosciences). Thrombin cleaved proteins were eluted from the GST gravity flow talon column in 2.5 ml and loaded directly onto the gel filtration column. Prior to loading, EDTA at a final concentration of 15 mM was added to the thrombin cleaved protein samples to bind any free calcium ions from the thrombin cleavage buffer. The ÄKTA prime (Amersham Pharmacia) FPLC was run at 3-4 ml/min for the HiLoad™ Superdex™ 200 and 75 columns and 1.5ml/min for the HiLoad™ Sephacryl™ S-100 column. Proteins were collected in 10 ml fractions. Protein containing fractions were confirmed by SDS PAGE and subsequently lyophilised.

Solutions:

Gel Filtration Buffer: 30 mM potassium phosphate, 100 mM NaCl (pH 7.0), filtered and degassed.

3.4.5 Desalting chromatography

Protein desalting was performed using 5 x 5ml HiTrap™ columns (Amersham Biosciences) pre-equilibrated with filtered and degassed Milli-Q water. The Milli-Q water was filtered with a 0.45 μm cellulose nitrate membrane (Whatman) and degassed before use. Lyophilised gel filtration samples were dissolved in 2 ml Milli-Q water and filtered using a 0.45 μM filter (Sartorius). The filtered sample was loaded onto the desalt column. The protein was eluted from the column first, followed by the NaCl. The protein containing fractions were confirmed by SDS PAGE. MAP-LC3 was unable to be desalted in water so the water was substituted for ammonium acetate which could also be freeze dried.

Solutions

Milli-Q water: filtered and degassed

Ammonium acetate: 25mM (pH 7), filtered and degassed

3.4.6 Lyophilising protein

Between the various columns attached to the AKTA prime and after the final desalt column protein containing fractions were concentrated by lyophilisation. The only exceptions to this were linear diubiquitin between cation exchange and gel filtration columns and MAP-LC3 between gel filtration and desalt columns. Proteins were snap frozen in liquid N_2 and applied to a micromodulyo freeze drier (Thermo Savant) for 24 - 48 hours or until all the liquid had been removed. MAP-LC3 was lyophilised in ammonium acetate after desalting and was additionally resuspended in water and re-lyophilised a further two times to remove any remaining ammonium acetate.

3.4.7 Spin concentrating protein

Where proteins could not be lyophilised, as is the case for the MAP-LC3 protein between gel filtration and desalt columns, proteins were spin concentrated using a Vivaspin20 spin concentrator (Vivascience). This spin concentrator had a 3000 Da molecular weight cut off Polyethersulfone (PES) membrance to remove small molecule and small protein contaminants.

3.4.8 Protein yield quantification

Desalted lyophilised protein was quantified by mass weight using a ALJ 120-4 balance (Kern). The desired amount of protein was weighed and dissolved in the appropriate buffer according to which experiments it was required for. Where protein could not be accurately weighed, concentration was calculated by measuring UV spectroscopy. A NanoDrop ND-100 spectrophotometer (NanoDrop) recorded the UV absorbance of the protein at 280 nm. The absorbance value was subsequently entered into the Beer-Lambert law.

Beer-Lambert law $A = \varepsilon c l$

Where A is absorbance (A_{280}), ε is the molar extinction co-efficient (M^{-1} cm^{-1}), c is concentration (M), l is path length (cm). $\varepsilon_{280\ nm}$ is calculated from the additive

contributions from Cys (125 M^{-1} cm^{-1}), Tyr (1490 $M^{-1}cm^{-1}$) and Trp (5500 M^{-1} cm^{-1}) residues[220].

3.5 Biophysical techniques

3.5.1 Electropray ionisation mass spectrometry (ESI-MS)

All electrospray ionisation mass spectrometry (ESI-MS) experiments were performed on a SYNAPT™ electrospray ionisation, high definition mass spectrometry (HDMS™) system with a Triwave™ ion mobility separation cell coupled to a quadrapole-time of flight (qTOF) mass analyser (Waters). A 100 µl syringe (Hamilton) was used to inject samples to the spectrometer at a rate of 5 µl min^{-1}. A standard electrospray source (Waters) or nanospray source (Waters) was attached to the mass spectrometer to transfer ions to the gas phase. For solvent exposed ESI-MS experiments, neat acetonitrile was placed in the electrospray source. Instrument control and initial data analysis was performed using the Masslynx™ software (Waters). Further analysis calculations were completed using Microsoft Excel 2007.

Solutions

ESI-MS buffer: 25mM Ammonium acetate

Nanospray ESI-MS buffer: 200mM Ammonium acetate

3.5.1.1 Native mass measurements

In order to confirm the identity of the protein or mutant native mass measurements were taken. Lyophilised protein was dissolved in ammonium acetate buffer to an approximate final concentration of 5 µM. ESI-MS was recorded in positive ion mode with an applied capillary voltage of 2.5 kV over 2-3 minutes. A sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323.15 K and trap and transfer collision energy voltages of 7.0 V and 6.0 V respectively and trap pressure of 2 mbar were also applied.

Data analysis:

The molecular mass was calculated from the mass to charge ratios observed in the mass spectrum using the mass lynx software. Each mass to charge ratio can be used to calculate the molecular mass using equation 1.

Equation 1 $$MWt = (MZ) - Z$$

Where MWt is the molecules molecular weight, M is the measured mass to charge ratio of the ion and Z is the charge state of that ion.

3.5.1.2 ESI-MS dilutions

ESI-MS dilution experiments were conducted in order to monitor the oligomerisation state of the p62 UBA mutants. The populations of monomeric and dimeric forms of the UBA vary with concentration, therefore all p62 UBA mutants were subjected to dilution experiments in a concentration range of 8 μM to 1 μM. ESI-MS was recorded in positive ion mode with an applied 2.5 kV capillary voltage over a time of 2-3 minutes. A sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323.15 K and trap and transfer collision energy voltages of 7.0 V and 6.0 V respectively and trap pressure of 2 mbar were applied.

Data analysis:

The apparent concentrations of monomer and dimer could not be accurately calculated. Where the mass and charge are both doubled an identical m/z value is recorded meaning peaks for the monomer and dimer overlap. Solvent exposed surface areas were therefore estimated using the average charge state[221]. The average charge state of the monomer was calculated using equation 2. The average charge state was subsequently plotted as a function of concentration, whereby changes in average charge state are easily visualised and used as an indicator of concentration dependent changes in monomer and dimer populations.

Equation 2 $$Z_{av} = (\sum ZI)/(\sum I)$$

Where Z_{av} is the average charge state for the species, Z is the charge state of a given ion of that species and I is the intensity of the given ion.

3.5.1.3 ESI-MS titrations

ESI-MS titrations were performed to measure binding affinities for the interaction between the S403 p62 UBA mutants (outlined in section 1.6) and ubiquitin. Moreover the stoichiometry of the newly formed complex could also be determined. Ubiquitin was titrated into mutant p62 UBA using the following concentrations of p62-UBA:ubiquitin (μM) 1:0.5, 1:0.75, 1:1, 1:1.25, 1:1.5 and 1:2. Competitive binding experiments containing all three S403 mutants were

mixed with ubiquitin (equimolar, 5 and 1 μM experiments). In order to measure the stoichiometry and confirm the formation of a ternary complex between ubiquitin, p62 and MAP-LC3 the three were mixed at the following concentrations of p62:ubiquitin:MAP-LC3 (μM) of 3:1.5:1.5 and 4:2:2. ESI-MS was recorded in positive ion mode with a 2.5 kV capillary voltage over a time of 2-3 minutes. A sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323.15 K and trap and transfer collision energy voltages of 7.0 V and 6.0 V respectively and trap pressure of 2 mbar were applied.

Data analysis:

The apparent concentrations of free protein ([P]), protein-ligand complex ([PL]) and free ligand (Protein 2) ([L]) were calculated from the concentrations of total protein and ligand used, and the signal intensities. Plots of [PL]/[P] vs. [L] yield a straight line of which the gradient is the association constant (K_a)[222]. Equation 3 is rearranged to give equation 4. This was subsequently used to calcluate the K_d using equation 5. The method relies on the assumption that complexes are transferred to the gas phase without significant perturbation to the forces responsible for their association. It also assumes that the intensities of ions recorded by the mass spectrometer are proportional to the concentrations of each of the species observed.

Equation 3 $$K_d = ([P][L])/[PL]$$

Where K_d is the dissociation constant, [P] is the concentration of protein 1, [L] is the concentration of ligand (or protein 2, as it is in this case) and [PL] equals the concentration of their complex.

Equation 4 $$[PL] / [P] = K_a [L]$$

Where K_a is the association constant, [P] is the concentration of protein 1, [L] is the concentration of ligand (or protein 2) and [PL] equals the concentration of their complex.

Equation 5 $$K_d = 1/K_a$$

Where K_d is the dissociation constant and K_a is the association constant.

3.5.1.4 nanospray ESI-MS

The GST fusion proteins were not able to be detected using the standard ESI-MS technique. The electrosrapy source was replaced with a nanospray source which is ideally suited to larger proteins. Using the standard nanospray source (Waters) a spectrum of the GST-p62 UBA construct was able to be recorded in order to confirm the molecular mass. Moreover titrations with monoubiquitin and linear diubiquitin were also recorded using this construct. ESI-MS was recorded in positive ion mode with a 1.5 kV capillary voltage over a time of 5 to 20 minutes. A sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323.15 K and trap and transfer collision energy voltages of 50 V and 30 V respectively and trap pressure of 2 mbar were applied.

3.5.2 Far-UV Circular dichroism (CD)

All far-UV Circular Dichroism (CD) experiments were recorded on an Pi-Star-180 Spectrophotometer (Applied Photophysics) interfaced with an Acorn Archimedes computer with inbuilt software (Applied Photophysics). The optical system was configured with a 75 W Xe lamp, circular light polarizer and end mounted photomultiplier. The temperature was regulated using an RTE-300 circulating programmable water bath (Neslab Inc) and a thermoelectric temperature controller (Melcor). Far UV-CD experiments were recorded using one of two quartz cuvettes. Higher protein concentrations (range of 1.5 – 0.1 mg/ml) utilised a 300 µl cuvette with a 1mm path length; whereas, lower protein concentrations (range of 0.15 – 0.01 mg/ml) a 3 ml cuvette with a 10 mm path length was used. All CD data was converted using the APL data converter version 3 (Applied Photophysics) and analysed using Microsoft Excel 2007.

Solutions

CD Buffer: 10 mM potassium phosphate (pH 7.0)

3.5.2.1 Far UV-CD scans

Far UV-CD scans were used to check the proteins had folded into the correct conformation and to confirm secondary structure. Scans were recorded in the wavelength range between 190 to 260 nm, data points were recorded at every 1 nm step. The number of counts was set to 10,000, adaptive sampling was enabled and set to 500,000 and entrance/exit slits of 3 nm were employed. All samples

were equilibrated at the appropriate temperature for 15 minutes prior to data collection.

Data analysis:

Ellipticity values recorded directly from the spectrophotometer were converted into molar ellipticity using equation 6 which corrects for wavelength and concentration[181].

Equation 6 $$[\theta]_{molar\lambda} = (100 \times \theta_{\lambda}) / (m \times d)$$

Where $[\theta]_{molar\lambda}$ equals the molar ellipticity, θ_{λ} is the observed ellipticity (degrees) at wavelength (λ), m is the protein concentration (decimolar) and d is the path length (cm). The units for molar ellipticity are deg cm^2 $dmol^{-1}$.

3.5.2.2 Thermal unfolding

Thermal unfolding studies were conducted in order to calculate the apparent midpoint (T_m) of unfolding. Far UV-CD monitored temperature melts were initially recorded between 278.15 K and 368.15 K using 0.5 K intervals. The temperature was controlled by a thermoelectric temperature controller (Melcor) with 0.2 K fluctuations from the specified temperature. The starting temperature was raised to 293.15 K when the recorded ellipticity of the previous melt remained constant between 278.15 K and 293.15 K. The ellipticity was recorded between 222 nm and 224 nm as this wavelength range represented the greatest difference between an ordered α helical structure and random coil.

Data analysis:

The apparent mid point of unfolding was calculated by using the first derivative of the folding curve as outlined by Greenfield *et al*[185]. This approach was used as an alternative to fitting the curve as the curve produced from the thermal unfolding profiles did not produce a defined end point.

3.5.2.3 Dilutions

Dilution experiments were recorded on the p62-UBA dimer mutants (outlined in section 1.6) in order to observe the presence of multiple species at equilibrium. CD spectra as well as thermal unfolding profiles were recorded at each point in the dilution The apparent T_m values were calculated from the thermal unfolding

data in order to detect any changes over the concentration range. Changes in T_m is indicative of multiple species at equilibrium. All mutants were monitored in the concentrations range 173 μM to 1 μM and were recorded by systematic 1:1 dilutions with CD Buffer. After dilution the samples were equilibrated for 15 minutes at room temperature before starting the next point in the dilution experiment. All concentrations were confirmed using a NanoDrop ND-100 spectrophotometer (NanoDrop).

Data analysis:

The apparent mid point was calculated at each concentration in the dilution by using the first derivative of the folding curve[185]. The apparent midpoint was plotted as a function of concentration to produce a concentration dependent apparent T_m graph. The T_m values for the dimer and monomer can be estimated from these graphs.

3.5.3 Isothermal titration calorimetry (ITC)

All isothermal titration calorimetry (ITC) experiments were performed on a VP-ITC high sensitivity titration calorimeter (MicroCal, Inc) at 298 K. All buffers and protein samples were degassed under vacuum for a minimum of 10 minutes before each experiment to remove any air bubbles and ensure efficient mixing and injection.

Solutions

ITC Buffer : 50 mM potassium phosphate, 50 mM NaCl (pH 7.0)

3.5.3.1 dilution ITC

Dilution ITC was used to monitor changes in heat upon dimer dissociation. Concentrated protein samples (250 μM – 300 μM) were titrated into a cell containing buffer alone. As the concentration of protein in the cell increases the heat change associated with dimer dissociation decreases producing a dissociation curve. The concentrated protein samples were prepared in phosphate buffer and degassed under a vacuum for a minimum of 10 minutes. Titrations consisted of 55 successive injections of 5 μl protein into buffer. A 300 second interval was permitted between injections. A constant stirring speed of 300 rpm ensured mixing during the titration. A reference power of 5 μCal/sec was used. Data were

analysed and fitted to a dimerisation dissociation model using Origin software (MicroCal).

Data analysis

The origin software calculates the dissociation affinity constant (K_{dim}) and the change in enthalpy (ΔH). These parameters were subsequently used to determine both the free energy of dissociation (ΔG) and the change in entropy of dissociation (ΔS) according to equations 7 and 8 respectively.

Equation 7 $$\Delta G = - R.T.\ln(K_{dim})$$

Where ΔG is the free energy of dissociation (KJ/mol), R is the molar gas constant (8.314 J mol^{-1} K^{-1}) and T is the temperature in Kelvin (298K) and K_{dim} is the dissociation affinity constant (μM).

Equation 8 $$\Delta G = \Delta H - T\Delta S$$

Where ΔG is the free energy of dissociation (KJ/mol), ΔH is the change in enthalpy (KJ/mol), T is temperature in Kelvin (298K) and ΔS is the change in entropy of dissociation (J/mol).

3.6 Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance experiments were performed on either an Avance™ III 600 MHz spectrometer (Bruker) fitted with a TXI triple resonance probe with Z-axis gradients (Bruker) or an Avance™ III 800 MHz spectrometer fitted with a Quadruple Resonance Inverse Detection QCI Cryoprobe, with Z-axis gradients. Solvent suppression was achieved using either WATERGATE[223] or excitation sculpting pulse sequences[224]. All spectra were acquired using standard Bruker pulse sequences and all samples were prepared in 90% H_2O and10 % D_2O. All experiments had transmitter frequencies set to water at 2823 and 3763 Hz for the 600 MHz and 800 MHz spectrometers respectively (4.7 ppm). The ^1H chemical shifts were referenced internally to the methyl peak of 4,4-dimethyl-4-silapentane-1-sulphonic acid (DSS) at 0.00 parts per million (ppm). ^{15}N and ^{13}C spectra were referenced indirectly using DSS and the IUPAC standard chemical shift ratios of 0.101 and 0.251 for ^{15}N:^1H and ^{13}C:^1H respectively[225]. Spectral processing was

carried out using Topspin 2.1 software. Protein assignments were achieved using CCPNMR software (version 2.1.2).

Data processing

All 2D and 3D spectra were phase corrected, baseline corrected, zero filled and an exponential window function applied to improve the signal to noise ratio and the resolution. The data was zero filled to 2048 in the indirect dimension(s), optimised with a shifted sine bell in all dimensions and an automatic polynomial baseline correction was applied. All processing was completed using Topspin 2.1.

3.6.1 Sample preparation

Lyophilised samples were dissolved in 600 µl of NMR buffer to final concentrations within the range of 10 µM – 1 mM. Low concentration samples (10 µM – 100 µM) could only be used on the highly sensitive 800 MHz spectrometer. Whereas high concentration samples (>100 µM) could be used on either the 600 MHz or 800 MHz spectrometers. Once dissolved the sample was centrifuged at 13, 000 rpm for 1 minute in a bench top centrifuge before transferring to a standard 5 mm NMR tube. For experiments conducted on the Bruker Avance 600 MHz spectrometer a 528-PP-7 NMR tube (Wilmad) was used. For experiments conducted on the Bruker Avance 800 MHz spectrometer a 535-PP-7 NMR tube (Wilmad) was used.

Solutions

NMR buffer: 25 mM potassium phosphate pH 7, 25 mM NaCl, 10% D_2O and 0.04% (w/v) sodium azide

3.6.2 1D NMR

1D NMR was used to check the protein had folded into the native state and the protein's suitability for further NMR experiments. Unlabelled protein samples of different concentrations (10 µM – 1 mM) were prepared. Spectra were recorded using a proton excitation sculpting pulse sequence[224]. The number of scans varied for different proteins, the concentration of the sample and the sensitivity of the machine being used; however, a range of 4 – 64 scans was normal. Spectral widths also varied but a range from 12 – 15 ppm was sufficient for all samples, with 12 ppm being used for the p62 UBA constructs and mutants.

3.6.3 2D NOESY and 2D TOCSY

The 2D homonuclear NMR experiments used in this thesis were the 2D NOESY[226] and 2D TOCSY[227]. These experiments were used to assign residues in the p62 UBA mutants which had perturbed significantly from those observed for the free form of the wild type p62 UBA. Unlabelled protein samples in the concentration range 250 μM to 1 mM were prepared. Spectra were acquired by collecting 2048 points in the direct dimension and 512 points in the indirect. Since this is a homonuclear experiment both dimensions are proton dimensions. A spectral width of 12 ppm was used in both dimensions. The number of scans was set to 64. The mixing times for the NOESY and TOCSY were 63.5 ms and 120 ms respectively.

Data analysis

The data obtained from the NOESY and TOCSY were used to assign residues which could not be assigned based on the ^{1}H-^{15}N amide chemical shifts reported for wild type. The TOCSY spectrum was used to identify the amino acid type as all amino acids have a pattern of cross peaks. The NOESY shows cross peaks between residues adjacent in sequence, therefore sequential assignment could be achieved. When the NOESY and TOCSY spectra are overlaid the peaks which belong to that particular amino acid and peaks which belong to nearby residues can be identified. These spectra are particularly useful for spectra that have been partially assigned.

3.6.4 ^{1}H-^{15}N HSQC and ^{1}H-^{15}N TROSY

The 2D heteronuclear NMR experiments used in this thesis were the ^{1}H-^{15}N HSQC[192, 193] and ^{1}H-^{15}N TROSY[194]. These experiments provide the correlation between the nitrogen and amide proton. All residues therefore produce an observable peak in the spectrum, with the exception of proline residues which lack an amide proton. For proteins less than 15 kDa the ^{1}H-^{15}N HSQC was used whereas proteins over 15kDa the ^{1}H-^{15}N TROSY experiment was utilized. The ^{1}H-^{15}N HSQC/^{1}H-^{15}N TROSY is recorded for the assignment of resonances so that each amide peak can be assigned to a particular residue in the protein. These experiments use the sensitivity of the backbone amide groups to study structural perturbations caused by mutation, construct extension and protein-protein

interactions. Using the amide chemical shifts previously reported for the wild type p62 UBA[140] resonances which have shifted in mutants or longer constructs can be identified. Titration studies were used extensively in this thesis to monitor the binding events between various p62 constructs and mutants to ubiquitin or linear diubiquitin. Dilution studies were also completed to monitor the monomer-dimer equilibrium of the p62 UBA mutants.

3.6.4.1 ^{1}H-^{15}N HSQC/^{1}H-^{15}N TROSY titrations

Titrations whereby an unlabelled protein is titrated into a ^{15}N labelled protein were used to probe protein-protein interactions. Experiments involving titrating unlabelled ubiquitin into ^{15}N labelled p62 UBA or a longer p62 construct (herein referred to as forward titrations) were recorded for T414A, T414K, T419K and S403D p62 UBA mutants as well as GST-p62 UBA, GST-hHR23A UBA2 and the longer GST-p62 constructs. Forward titrations whereby linear diubiquitin was titrated in were performed for the GST-p62 UBA and GST-hHR23A UBA2. Depending on the size of the complex formed either the ^{1}H-^{15}N HSQC or ^{1}H-^{15}N TROSY was recorded for the free and fully bound samples as well as each point in the titration. A 1D spectrum was recorded before each ^{1}H-^{15}N HSQC or ^{1}H-^{15}N TROSY experiment.

A 1200 µl sample of ^{15}N labelled protein was prepared and split into two 600 µl samples. One sample was used as the unbound sample. Lyophilised unlabelled protein was added to the other sample at a 4:1 molar ratio to form a fully bound sample. If the titration was completed on the 600 MHz spectrometer a 1 mM sample was split into two 500 µM samples and 2mM ubiquitin was added. If the titration was completed on the 800 MHz spectrometer a 500 µM sample was split into two 250 µM samples and 1mM ubiquitin was added. The two samples were mixed in various quantities according to equation 9 to generate a range of concentrations of unlabelled protein. A specific volume was removed from the free sample and replaced with the same volume from the fully bound sample. This way the 600 µl volume was maintained whilst titrating in the ubiquitin or diubiquitin.

Equation 9 $V2 = ((M1-M3)*V1)/(M1-M2)$

Where V2 equals the volume which is removed from the unbound sample. This volume is replaced by the fully bound sample. M1 is the concentration of the current sample, M3 is the desired concentration, V1 is the volume of the sample and M2 is the concentration of the concentrated stock.

For the intermediate points in the forward titrations 2048 points and 64 points were acquired in the proton and ^{15}N dimensions respectively. A spectral width of 13 ppm was used for the proton dimension. A spectral width of 40 ppm was used for the ^{15}N dimension. The number of scans was set between 32 and 64 depending on the protein used. For the free and fully bound samples a higher resolution spectrum was recorded whereby the number of points in the nitrogen dimension and the number of scans were increased.

Titrations whereby ^{15}N-ubiquitin was titrated into unlabelled p62 UBA (herein referred to as reverse titrations) were also recorded. Reverse titrations were recorded for the S403A, S403D and S403E p62 UBA mutants as well as GST-p62 UBA. A reverse titration whereby unlabelled GST-p62 UBA was titrated into ^{15}N linear diubiquitin was also completed. The reverse titrations also used the method outlined in equation 9. For the points in the reverse titrations 2048 points and 64 points were acquired in the proton and ^{15}N dimensions respectively. A spectral width of 13 ppm was used for the proton dimension. A spectral width of 40 ppm was used for the ^{15}N dimension. The number of scans was set to 32. For the free and fully bound samples a higher resolution spectrum was recorded whereby the number of points in the ^{15}N dimension and the number of scans were increased.

Data analysis

Changes in the chemical shifts of amino acids were used to identify binding patches and determine binding affinities. Chemical shift perturbations were calculated using equation 10.

Equation 10 $$CSP = \sqrt{((\Delta^1H)^2 + (\Delta^{15}N/5)^2)}$$

Where Δ^1H and $\Delta^{15}N$ are the observed changes in proton and ^{15}N chemical shift in response to binding.

Reverse titrations were used to calculate an apparent K_d for the interaction. Final CSP values greater than 0.05 ppm were entered into IGOR Pro Version 5.0.57 (Wavemtrics) and fitted using equation 11.

Equation 11 $$\delta = \Delta\delta\ ([L]\ /(K_d + [L])$$

Where δ is the observed CSP at a given ligand concentration, Δδ is the final CSP at saturation and [L] is the unlabelled protein concentration.

K_d values whereby the competitive processes of dimerisation and binding were taken into account were calculated according to equation 12. The K_d values produced from fitting in IGOR Pro is used as the K_{obs} and K_{dim} values were calculated from ITC dilution experiments.

Equation 12 $$K_{obs} = K_d^2/K_{dim}$$

3.6.4.2 ^{1}H-^{15}N TROSY titrations to form a p62 mediated ternary complex

For the titrations to form the p62 mediated ternary complex a different approach was used. A 250 μM sample of the C331S mutant of the p62 300-440 subfragment was prepared by dissolving lyophilized protein in 600 μl NMR buffer. MAP-LC3 was added to the C331S p62 300-440 followed by ubiquitin and vice versa. A concentrated stock of ubiquitin was used to spike in various concentrations up to 1 mM (4:1 molar ratio). The MAP-LC3 was insoluble above 250 μM so a concentrated stock could not be spiked in. Therefore a 250 μM (1:1 molar ratio) sample was prepared by dissolving lyophilized protein in water and splitting into aliquots. The aliquots were subsequently free dried and added to the free C331S p62 300-440 or bound to ubiquitin complex NMR sample. The interaction of the p62 with ubiquitin (via the UBA) and MAP-LC3 (via the LIR) have been previously reported separately, providing indications into the concentrations of these proteins required for saturation of the p62 construct.

All spectra within the titration were recorded with 2048 points in the proton dimension and 128 points in the ^{15}N dimension. Spectral widths of 16 ppm in the proton dimension and 35 ppm in the ^{15}N dimension were applied. The number of scans used for each point in the titration was just 4 scans due to the highly sensitive nature of the C331S p62 300-440.

3.6.4.3 ^{1}H-^{15}N HSQC/TROSY dilutions

Wild type p62 UBA, the T419K p62 UBA mutant and the GST-T419K p62 UBA were diluted from either 500 μM (for T419K) or 250 μM (for wild type p62 UBA and GST-T419K). The starting concentration was reduced by 50 μM and a spectrum recorded. This was repeated until the concentration was 50 μM. After this concentration systematic 1:1 dilutions with NMR buffer was used to make concentrations of 25 μM, 12 μM and 6 μM (6 μM was only used in the wild type p62 UBA dilution). Wild type data was recorded as a control whereas, the T419K and GST-T419K were recorded to observe what concentration the dimeric species is no longer observable. The p62 UBA mutants T414A and T414K were not systematically diluted. Spectra for a high concentration sample (500 μM) and a low concentration sample (10 μM) were recorded only due to the highly sensitive nature of these mutants. This high and low concentration approach was also applied to the p62 UBA mutants S403A, S403D and S403E in the interest of saving time. Thus 12 μl of the 500 μM sample was added to 488 μl NMR buffer to make the 10 μM sample.

The wild type p62 UBA and all p62 UBA mutants used ^{1}H-^{15}N HSQC; whereas, the GST-T419K used the ^{1}H-^{15}N TROSY. The spectra were recorded with 2048 points and 64 points in the proton and ^{15}N dimensions respectively. Spectral widths were set to 12 ppm in the proton dimension and 35 ppm in the ^{15}N dimension. The number of scans increased when the concentration was decreased over the course of the dilution. As a general rule when the sample concentration was halved the number of scan should be multiplied by four to get the same quality spectrum.

3.6.4.4 ^{15}N Heteronuclear NOE (^{15}N (^{1}H) NOE)

The ^{15}N heteronuclear NOE (^{15}N (^{1}H) NOE) was measured using a ^{15}N-HSQC based pulse sequences with and without proton saturation during the relaxation delay prior to the starting 90° ^{15}N pulse[195]. The experiments were run as interleaved ^{15}N HSQCs which were subsequently split into a pair of ^{15}N HSQCs, one with proton saturation and one without.

An experiment was performed on the T419K mutant at a concentration of 25 μM in order to observe the flexibility of the monomeric form of this mutant. This

experiment was performed by collecting 2048 and 512 points in the proton and ^{15}N dimensions respectively. Spectral widths of 13 ppm for the proton dimension and 35 ppm for the ^{15}N dimension were also employed. The number of scans was set to 52.

Experiments were also performed on the C331S p62 300-440 bound to either monoubiquitin or MAP-LC3 or both proteins as part of a ternary complex. These experiments were performed by collecting 2048 and 256 points in the proton and ^{15}N dimensions respectively. Spectral widths of 16 ppm for the proton dimension and 35 ppm for the ^{15}N dimension were also employed. The number of scans was set to 40.

Data analysis

The interleaved data was separated into two HSQC spectrums, one with and on without proton saturation using the split2 command in Topspin 2.1. The $^{15}N(^{1}H)$ NOE values were calculated using Excel 2007 using equation 13.

Equation 13 $$^{15}N(^{1}H)\ NOE = (I_{sat}/I_{eqm})$$

Where I_{sat} is the intensity of of a particular peak with proton saturation and I_{eqm} is the intensity of the same peak without proton saturation.

3.6.5 3D ^{15}N-HSQC-NOESY and 3D ^{15}N-HSQC-TOCSY

The 3D double resonance NMR experiments used in this thesis were the ^{15}N-HSQC-NOESY[193,192,228,229] and ^{15}N-HSQC-TOCSY[193,192,230]. These spectra contain two proton dimensions and one ^{15}N dimension. These experiments were used to assign residues in the p62 UBA mutants which had perturbed significantly from those observed for the bound form of the wild type p62 UBA. A fully bound sample was prepared using a 4:1 molar ratio of ubiquitin:UBA.

Spectra were acquired by collecting 2048 points in the direct proton dimension over a spectral width of 12 ppm. 128 increments per plane in the indirect proton dimension and 32 increments per plane for the ^{15}N dimension were recorded. Spectral widths of 12 ppm for the indirect proton and 35 ppm for the ^{15}N dimensions were applied. The number of scans was set to16. The mixing time was 200 ms and 60 ms for the ^{15}N-HSQC-NOESY and ^{15}N-HSQC-TOCSY

respectively. The quality of the HSQC-TOCSY could have been vastly improved by using a longer mixing time (120 ms to 200 ms).

Data analysis

The data obtained from the ^{15}N-HSQC-NOESY and ^{15}N-HSQC-TOCSY were used to assign resonances which had deviated from the chemical shifts reported for the wild type p62 UBA. The data was analysed using a similar approach to the 2D NOESY and 2D TOCSY, whereby the amino acid type could be identified using the ^{15}N-HSQC-TOCSY and the ^{15}N-HSQC-NOESY is subsequently used to connect the amino acids in a sequence.

3.6.6 3D triple resonance NMR

Standard back bone assignment experiments for the assignment of protein NH, HN, C', Cα and Cβ chemical shifts were performed on a free 250 μM ^{15}N-^{13}C labelled C331S p62 300-440 sample and bound to MAP-LC3 (250 μM) and ubiquitin (1 mM) forming a ternary complex. The 3D triple resonance experiments were recorded to assign the p62 LIR and any UBA peaks which have shifted in the free or bound forms. The NHCACB[206], HN(CO)CACB[206] and HNCA[208] experiments were used to identify the Cα and Cβ carbons whereas the HNCO[208],[231] and HN(CA)CO[208] experiments were used to identify the C' carbons.

For the HNCACB and HN(CO)CACB experiments 1024 points were acquired in the proton dimension. Also 32 increments per plane in the indirect ^{15}N dimension 128 increments per plane in the indirect ^{13}C dimension were recorded. The spectral widths were set to 12, 35 and 75 ppm for the proton, nitrogen and carbon dimensions respectively. The number of scans was set to 24 per increment.

For the HNCA experiment 1024 were acquired in the proton dimension. Also 32 and 64 points per increment were recorded in the indirect ^{15}N and ^{13}C dimensions respectively. The spectral widths were set to 12, 35 and 32 ppm for the proton, ^{15}N and ^{13}C dimensions respectively. The number of scans was set to 32 per increment.

For the HNCO and HN(CA)CO experiments 1024 points were acquired in the proton dimension. Also 32 increments per plane in the indirect ^{15}N dimension and 64 increments per plane in the indirect ^{13}C dimension were recorded. The spectral

widths were set to 12, 35 and 25 ppm for the proton, ^{15}N and ^{13}C dimensions respectively. The number of scans was set to 32 per increment.

Data analysis

The spectra were processed using topspin 2.1. Copies of the processed data were made and the copies imported into CCPNMR analysis software (version 2.1.2). Standard protein backbone assignment methodologies were used employed to assign the p62 LIR (in the free and bound forms) and p62 UBA (bound form).

4.0 Purification of monoubiquitin and linear diubiquitin

4.1 Introduction

Results chapters 6 to 9 involve the binding of p62 constructs containing the UBA domain or p62 UBA mutants to ubiquitin ligands. For the purpose of these studies it was therefore necessary to engineer both monoubiquitin and linear diubiquitin. Monoubiquitin (*S. Cerevisiae*) was previously cloned by Dr Geoff Platt (formerly of the Searle group) into the pKK223-3 plasmid between the *EcoRI* and *HindIII* restriction sites in the multiple cloning site. The plasmid was purchased from Amersham Pharmacia and contains ampicillin resistance. Linear diubiquitin (*Homo Sapien*) was received pre-cloned into the pt7-7 plasmid between the restriction sites *NdeI* and *EcoRI* as a gift from Paul Sheppard (formerly ENZO life Sciences). The pt7-7 plasmid also contains ampicillin resistance. Despite using ubiquitin from different organisms, the difference between the ubiquitin orthologues is only three residues, with the tertiary structures shown to be virtually identical.

Previous work, as part of a collaborative project between the Searle and Layfield groups (both University of Nottingham), have shown that the p62 UBA shows an enhanced binding affinity for polyubiquitin chains over monoubiquitin in pull down assays using immobilised p62[140]. In order to monitor the increased affinity as potentially a result of the avidity effect, monoubiquitin binding was compared to diubiquitin binding (chapter 9). Of the eight different types of diubiquitin, linear diubiquitin was chosen for these binding studies due to its structural similarity to Lys63 linked diubiquitin[35]. The p62 protein is involved in the NF-κB signalling pathway, a pathway that is known to be regulated by both linear[74,76,77] and Lys63 linked polyubiquitin chains[61]. Lys63 is located very close to the C terminal causing structures that differ subtly in the linker region, rather than in the orientation or compactness of the subunits. Many UBDs have been shown to preferentially bind Lys63 and linear chains over Lys48 chains[232,233], emphasising their structural similarities. Linear chains can be expressed and purified by cloning one ubiquitin moiety after the other into plasmids, meaning they have the advantage of not requiring specific ubiquitination enzymes as well as being a more financially viable.

4.2 Results

4.2.1 Protein overexpression

Both monoubiquitin and linear diubiquitin DNA were purified from XL-1 blue *E.coli*. The DNA was transformed by heat shock into C41 (DE3) *E.coli* and plated onto agar plates containing ampicillin only. The C41 (DE3) strain was used because, like the BL21 (DE3) strain it was modified from[217], it is deficient in *Ion* and *ompT* proteases which allows for better isolation of intact recombinant proteins, but unlike the BL21 (DE3) strain it is able to express toxic proteins. C41 (DE3) *E.coli* is a cell line which contains the T7 polymerase under the control of the IPTG inducible *tac* promoter. The hybrid *tac* promoter contains the *lac* promoter sequence, enabling the use of IPTG to initiate protein overexpression. IPTG, a β galactosidase analogue, permits T7 polymerase activity by binding to the *lac* repressor at the *lac* promoter sequence releasing the repressor from the DNA allowing the plasmid gene of interest to be expressed.

Small scale protein overexpression trials were completed in 10ml universals. After confirmation of protein expression by SDS PAGE, large scale growths ensued. Overnight cultures were grown and used to inoculate media (1 x 20 ml or 3 x 15 ml cultures for 1 litre of luria broth or M9 minimal media respectively). Ampicillin was added at inoculation to select for plasmid containing cells which possess antibiotic resistance. Monoubiquitin was expressed in M9 minimal media as both an unlabelled and ^{15}N labelled protein to increase the purity of the expression and prevent overgrowth of cells. Ubiquitin is an extremely stable and tolerant protein, it was not unsurprising overgrowth of cells could occur in a nutrient rich media. Linear diubiquitin showed poor expression in M9 minimal media but expressed well in luria broth. Luria broth media was therefore chosen for large scale overexpression of linear diubiquitin as an unlabelled protein. M9 minimal media was utilised for ^{15}N labelling of linear diubiquitin. The cells of the inoculated media were grown at 37 °C until the OD_{595} reached 0.7, a density where the cells have reached an exponential growth rate. The majority of cells are at their optimum for protein expression at OD_{595} 0.7. Subsequently overexpression was induced by the addition of IPTG. Over time the ampicillin was metabolised by the *E. coli* cells, therefore further ampicillin was added to maintain the selection for plasmid containing cells only. The incubation temperature was

lowered to 30 °C for both monoubiquitin and linear diubiquitin at induction. Monoubiquitin was expressed for 16 hours whereas linear diubiquitin was expressed for 8 hours as an unlabelled protein and 16 hours as a ^{15}N-labelled protein. The duration of expression reflects the difference between the two different medias used for growth. The cells were harvested in a centrifuge once the OD_{595} could not get any higher. The cell pellets were stored at -80 °C until needed for purification. Gel samples were taken prior to protein induction and cell harvesting. The gel samples were visualised on a 20 % acrylamide gel. This percentage of acrylamide was used due to the small molecular weights of monoubiquitin and linear diubiquitin. The expression of monoubiquitin is shown in figure 4.1 whereas the expression of linear diubiquitin is included in figure 4.2 The difference in the intensity of the bands corresponding to the overexpressed monoubiquitin and linear diubiquitin reflects the media used for the overexpression.

.

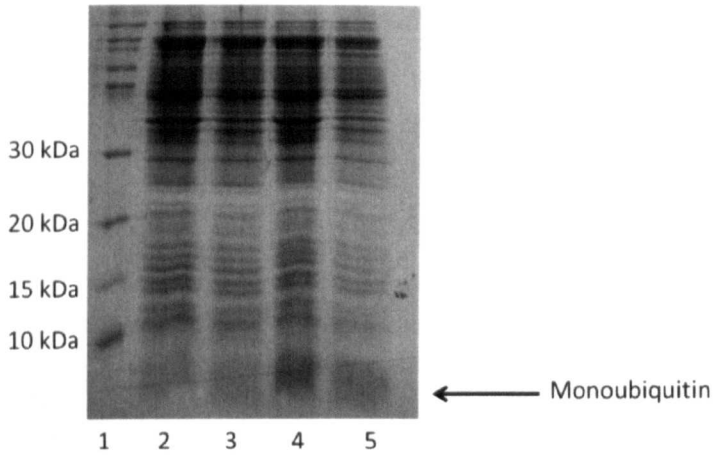


Figure 4.1 – A 20 % acrylamide SDS PAGE gel of the overexpression of monoubiquitin in C41 (DE3) *E.coli* in M9 minimal media. A Novex prestained protein standard marker (Invitrogen) was used as a size standard (lane 1). The total and soluble fractions of cell lysate prior to protein induction with IPTG are shown in lanes 2 and 3 respectively. The total and soluble fractions of cell lysate 16 hours after the addition of IPTG are depicted in lanes 4 and 5 respectively.

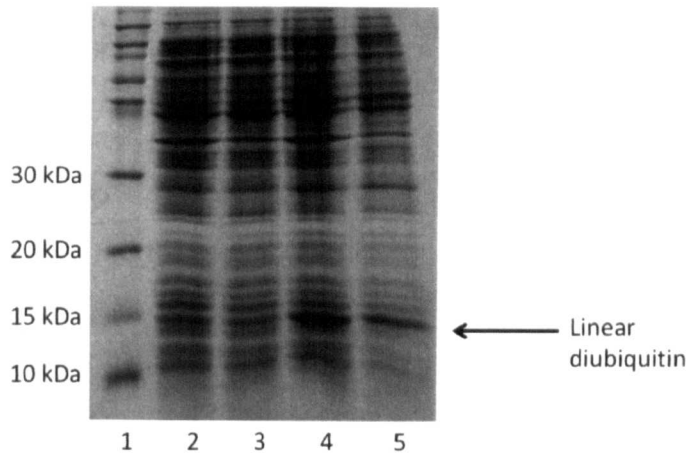


Figure 4.2 – A 20 % acrylamide SDS PAGE gel of the overexpression of linear diubiquitin in C41 (DE3) *E.coli* in luria broth nutrient rich media. A Novex prestained protein standard marker (Invitrogen) was used as a size standard (lane 1). The total and soluble fractions of cell lysate prior to protein induction with IPTG are shown in lanes 2 and 3 respectively. The total and soluble fractions of cell lysate 8 hours after the addition of IPTG are depicted in lanes 4 and 5 respectively.

4.2.2 Protein purification

Despite different overexpression conditions, both monoubiquitin and linear diubiquitin were both purified using a virtually identical purification protocol. Three purification columns were utilised for this, a cation exchange column, a gel filtration column and a desalt column. Confirmation of protein containing fractions was achieved by visualising on an SDS PAGE gel (figures 4.3 and 4.4).

Cell pellets were removed from -80 °C storage and allowed to thaw in a minimal volume of 5 x AE buffer and DNase I (0.01 mg/ml). Further 5 x AE buffer was added to drag the pH of monoubiquitin or linear diubiquitin to pH 4, where several other proteins are insoluble. The cells were subsequently lysed by sonication, releasing the overexpressed protein into solution. Two rounds of high speed centrifugation removed the unwanted cellular debris and insoluble protetin content. The ubiquitin containing supernatant was diluted to 150 ml with degassed Milli-Q water, filtered and degassed and loaded onto a pre-equilibrated cation exchange column attached to an AKTA prime FPLC system. Monoubiquitin and linear diubiquitin have a pI value of approximately 6.79, making them positively charged at pH 4. All positively charged proteins are capable of binding to the negatively charged resin of the column. The protein was then eluted from the

column by applying a linear NaCl gradient whilst the absorbance at 280 nm is recorded. The gradients applied to the proteins differed due to the presence of extra proteins expressed in the luria broth media the linear diubiquitin was grown in. Monoubiquitin was eluted using a sharp gradient of 0 -100 % over 100 ml. A shallower gradient of 0 – 60 % over 500 ml was applied to elute linear diubiquitin. A contaminant protein with a similar molecular weight to linear diubiquitin meant that altering the elution gradient on the cation exchange column was the only way of removing the contaminant. The protein was collected in both cases as 10 ml fractions. Monoubiquitin fractions were pooled and snap frozen in liquid N_2 prior to 48 hours on the freeze drier. Linear diubiquitin showed some difficulties on the freeze drier, probably due to the increased NaCl content. Linear diubiquitin was therefore applied directly to the next column, the gel filtration column, as the 10 ml fractions collected from the cation exchange column.

Gel filtration was used to remove any non-specific binders from the fractions collected from the cation exchange column. Both monoubiquitin and linear diubiquitin utilised a Superdex 75 gel filtration column. Freeze dried monoubiquitin was diluted to 2.5 ml with phosphate buffer and loaded onto the gel filtration column, whereas linear diubiquitin was loaded as 10 ml fractions from the cation exchange column using a 10 ml superloop. A peak corresponding to the protein is recorded by monitoring the absorbance at 280 nm. Monoubiquitin typically eluted after 280 – 300 ml and linear diubiquitin after 260 – 300 ml. All gel filtration experiments were conducted in phosphate buffer to ensure any contaminant acetate signals had been completely removed ensuring suitability for NMR experiments. Again the proteins were collected as 10 ml fractions, with the protein containing fractions being snap frozen in liquid N_2 prior to freeze drying for 48 hours. Linear diubiquitin was more dilute after gel filtration than the cation exchange column. The reduced salt content permitted freeze drying the fractions.

The final step of the purification was a desalt column. The desalt column, like the gel filtration column separates molecules according to their size and shape. The protein is therefore eluted before the NaCl bound to the protein to enhance its stability. Removing excess NaCl ensures the protein is viable for biophysical analysis. Lyophilised monoubiquitin and linear diubiquitin fractions from the gel filtration column were diluted to 2 ml with Milli-Q water and loaded onto a 5 ml

HiTrap desalt column. A peak corresponding to the protein is recorded by monitoring the absorbance at 280 nm. Both proteins typically eluted after 7 ml and were snap frozen in liquid N_2 and freeze dried for 24 hours. Approximately 10 - 20 mg pure protein was produced from a 1 litre growth for both proteins.

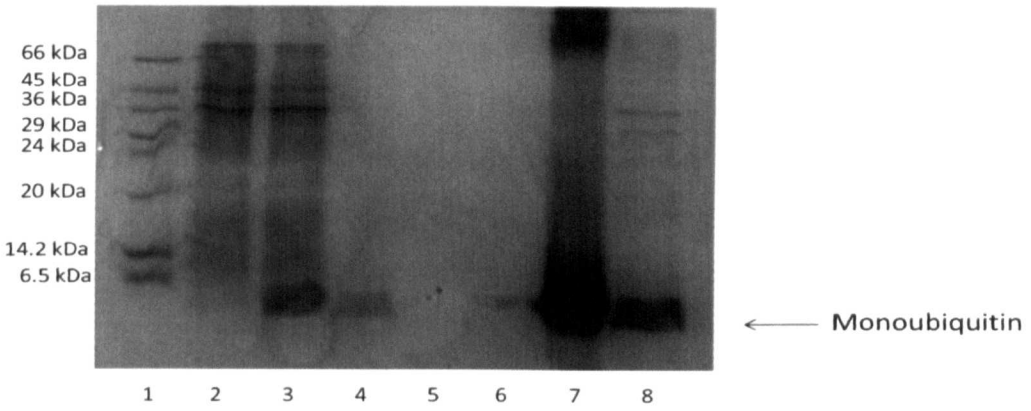


Figure 4.3 - A 20 % acrylamide SDS PAGE gel of a cation exchange column of monoubiquitin expressed in C41 (DE3) *E.coli*. A low molecular weight marker (Sigma) was used as a size standard (lane 1). Samples were taken prior to protein induction with IPTG, 16 hours after the addition of IPTG and the soluble cell lysate fraction are depicted in lanes 2-4 respectively. Cation exchange fractions 2, 21, 22 and 23 are present in lanes 5-8. Fraction 2 was used as a control. Fractions 22 and 23 were shown to be the monoubiquitin containing fractions however both fractions contain some impurities. The fractions were snap frozen in liquid N_2 separately and freeze dried for subsequent purifications.

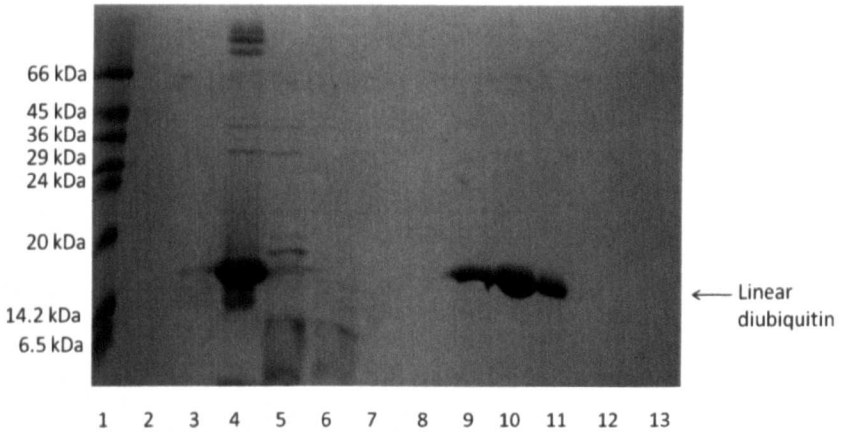


Figure 4.4 - A 20 % SDS PAGE gel of a cation exchange column and gel filtration fractions of linear diubiquitin expressed in C41 (DE3) *E.coli*. A low molecular weight marker (Sigma) was used as a size standard (lane 1). Cation exchange fractions 1-5 reflecting the increasing the NaCl gradient used to elute from the column are shown in lanes 2-6. Fraction 3 from the cation exchange column (lane 4 in this gel) was directly loaded onto a gel filtration column. Gel filtration fractions corresponding to fractions 25-31 are depicted in lanes 7-13. The protein containing fractions (27-29) were pooled for desalting.

5.0 Purification of cleaved and uncleaved proteins cloned into pGEX plasmids

5.1 Introduction

Various p62 constructs, hHR23A UBA2 and the MAP-LC3 were expressed and purified for subsequent study in chapters 6 to 9 of this thesis (table 5.1). All proteins that were purified from these plasmids utilised a very similar purification protocol as both plasmids possess an N-terminal GST tag, have ampicillin resistance and are under the control of a *tac* promoter. A *tac* promoter is a hybrid promoter derived from sequences of the trp and lac UV5 promoters permitting higher efficiency transcription of genes. The tac promoter is repressed by the lac repressor and activated with the β galactosidase analogue IPTG. The subtle differences in the purification procedures were attributed to the different intrinsic properties of each protein or mutant. The proteins purified in chapters 6 to 9 can be broadly divided into four categories, (i) p62 UBA helix 2 dimerisation interface mutants, (ii) p62 UBA phosphomimetic mutants, (iii) those required for the formation of a p62 mediated ternary complex and (iv) GST tagged proteins used to model the full length p62 protein. The GST fusion of the hHR23A UBA2 protein was used as a control where appropriate. This domain has been studied extensively and has been confirmed to exist as a monomer[95,96].

Table 5.1 – The various proteins cloned into pGEX plasmids and used throughout this study. The restriction sites indicate those which the gene of interest were cloned between. The tag found within the plasmid and the extension which remains after thrombin cleavage are also indicated for each protein.

Protein	Plasmid	Restriction sites	Affinity tag	Extension after cleavage
P62 387-436 (UBA)	pGEX-4T-1	BamHI-XhoI	GST	GS
P62 261-440	pGEX-4T-1	EcoRI-XhoI	GST	GSPEF
P62 300-440	pGEX-4T-1	BamHI-XhoI	GST	GS
P62 341-440	pGEX-4T-1	EcoRI-XhoI	GST	GSPEF
hHR23A UBA2	pGEX-4T-1	BamHI-XhoI	GST	GS
MAP-LC3	pGEX-4T-3	EcoRI-SalI	GST	GSPNS

5.2 Results

5.2.1 Cloning

The different p62 constructs and hHR23A UBA2 were all cloned into the vector PGEX-4T-1. The hHR23A UBA2, the p62 UBA domain (corresponding to residues 387-436) and the p62 subfragment corresponding to residues 341-440 were previously cloned by the Layfield group. The MAP-LC3 protein was also cloned by the Layfield group but into the pGEX-4T-3 plasmid.

Fragments of the p62 corresponding to residues 300-440 and 261-440 were personally cloned into the multiple cloning site of the pGEX-4T-1 plasmid between the *BamHI* and *XhoI* sites and the *EcoRI* and *XhoI* sites respectively. The p62 subfragments were amplified by PCR and the product purified from enzymatic material using a gel extraction kit. Equivalent restriction digest reactions for both the plasmid and the amplified PCR insert were set up in thin walled PCR tubes and heated at 37 °C until reaction completion. The restriction endonucleases in these reactions produced complementary staggered ends by cutting double stranded DNA at a specific sequence. Both the plasmid and insert were purified from other components of the restriction digest reaction by further gel extraction. The plasmid and the insert were subsequently ligated together using a T4 DNA ligase which functions to repair the breaks within DNA sequences. The ligated DNA was transformed into XL1-blue *E.coli* and plated onto agar plates containing ampicillin and tetracycline antibiotics. Ampillicin was used to select for the plasmid whereas tetracycline was used to select for XL1-blue. The DNA produced by the *E.coli* was purified using a DNA miniprep kit and sent for DNA sequencing[214] to confirm the presence and integrity of the insert within the plasmid.

5.2.2 Site directed mutagenesis

The site directed mutagenesis reactions for the helix 2 Ala mutants, the C331S p62 300-440 mutant and the S403E mutant were completed previously by Dr Jed Long, Miss Jill Sollenberger and Miss Alice Goode respectively (all of the Searle group, University of Nottingham).

Complementary forward and reverse primers, between 30 and 40 base pairs, were designed based on the *E.coli* tRNA codon usage and used for the production of the helix 2 Lys mutants and the Ser403A and S403D mutants. The desired mutation was located in the centre of the primers (see primer design in appendix), with each end of the primers terminating in one or more G or C bases. These bases are able to form an extra hydrogen bond with the template DNA providing a more stable interaction between the two strands. PCR mutagenesis reactions were performed on a thermal cycler. The program consisted of denaturing, annealing and extension steps, which were cycled 16 times to incorporate the single amino acid change. The double helix of the parental DNA was denatured allowing the primers to anneal to the correct positions within the separated strands using Watson and Crick base pairing. The primers were elongated by *Pfu turbo* DNA polymerase to produce complementary strands containing the desired mutation. Upon reaction completion Dpn1 endonuclease was added to the reaction to digest parental DNA. This enzyme is selective for methylated DNA found in genomes. Therefore any plasmids containing either one or both strands as methylated DNA will be digested leaving behind only newly mutated DNA. Gel samples were taken before and after the addition of Dpn1 and visualised on a 1 % (w/v) agarose gel (figure 5.1).

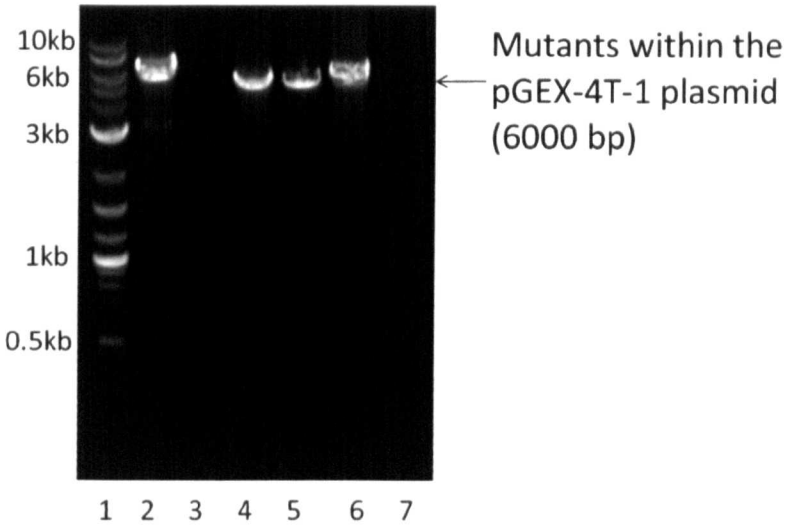


Figure 5.1 – 1 % (w/v) agarose gel of the PCR product from the site directed mutation reaction of the T414K, R415K, Q418K and T419K mutants A 2-log ladder (lane 1) was used as a size standard reference. Lanes 2 and 3 are other mutagenesis reactions completed at the same time as the lysine mutants. Lanes 4, 5,6 and 7 correspond to T414K, R415K, Q418K and T419K respectively. All samples were taken after the addition of Dpn1 restriction enzyme and therefore respresent newly mutated DNA. The T419K reaction was unsuccessful on this attempt.

The PCR product was subsequently transformed into supercompetent XL1-blue and plated onto agar plates containing both ampicillin (selective for the desired plasmid containing the PCR product) and tetracycline (selective for XL1-blue *E.coli*). Heat shock was used for plasmid DNA uptake into the host system. A 10 ml overnight culture containing the antibiotic selections outlined above was inoculated with a single isolated colony grown on the agar plate. The DNA produced by the XL1-blue in the overnight culture was purified using a miniprep kit and sequenced using the standard Sanger method to verify the mutation. The XL1-blue strain of *E.coli* was chosen because of their ability to produce high quality DNA from a miniprep kit. XL1-blue are endonuclease (*endA*) and recombination (*recA*) deficient which improve the purity of the DNA and increase the stability of the inserted DNA within the plasmid respectively.

5.2.3 Protein overexpression

All of the cleaved proteins, namely the p62 UBA mutants, the C331S 300-440 construct and MAP-LC3 protein were purified using the same methodology as that used for wild type p62 UBA, but with minor alterations to suit each individual protein or mutant. The purified DNA from XL1 blue *E.coli* was transformed using the heat shock methodology to C41 (DE3) *E.coli* and plated onto agar plates containing ampicillin only. The C41 (DE3) strain is deficient in *Ion* and *ompT* proteases, which allows effective expression of recombinant proteins from all classes of organisms. The C41 (DE3) strain also contains at least one mutation compared to its parental BL21 (DE3) strain which permits the expression of proteins that are toxic to other cell lines[217]. Colonies were grown on ampicillin resistant agar plates were used to make the glycerol stocks required for further protein production.

Large scale protein overexpressions were completed in 2 litre conical flasks after confirmation of protein expression had been achieved in small scale tests. The Q418K mutant was unable to be expressed in either BL21 (DE3) or C41 (DE3) strains of *E.coli*, it was therefore deemed a lethal mutation. A 20 ml overnight culture was grown and used to inoculate luria broth for unlabelled proteins. M9 minimal media was used for the production of ^{15}N and ^{13}C-^{15}N labelled proteins and 3 x 15 ml overnights were used to inoculate this media. The cells were grown at 37 °C until the OD_{595} reached 0.7, the optimum density for protein induction. The temperature was cooled to 30°C at OD_{595} 0.7 and protein overexpression was induced by the addition of IPTG. Supplementary ampicillin was added to maintain the selection for plasmid containing cells. The cells were induced for the 4 to 16 hours, depending on the protein and whether it was isotopically labelled. Finally the cells were harvested when a higher OD reading could not be obtained. In both small and large scale overexpressions gel samples were collected before and after protein induction and visualised on a 20 % acrylamide SDS PAGE gel (figures 5.2 and 5.3). In the pGEX-4T-1 plasmid a second unknown band appears at around 29 kDa after induction with IPTG. In 20 % acrylamide SDS PAGE gels the GST-p62 UBA or mutant UBA expression band is just above the expression band corresponding to the unknown protein. When expressing longer GST p62 constructs there is greater separation between these two bands.

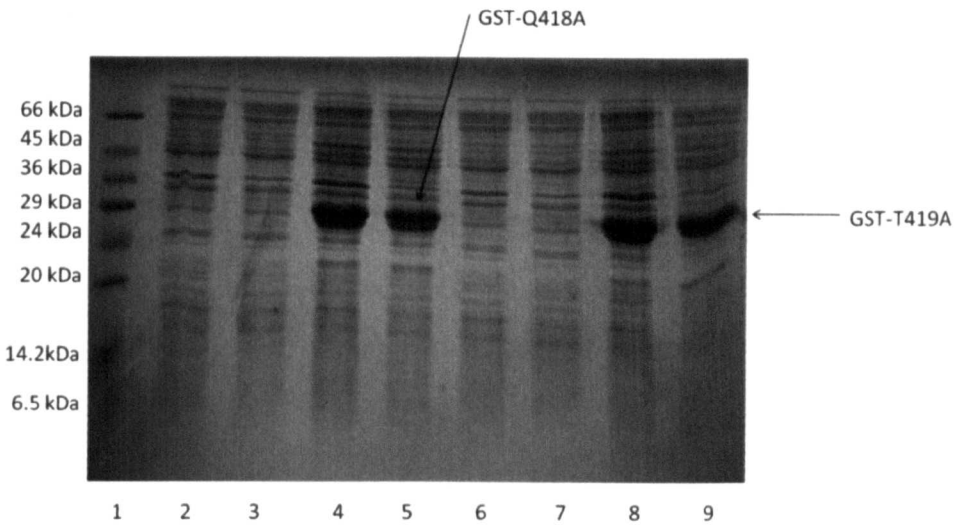


Figure 5.2 - A 20 % acrylamide SDS PAGE gel showing the overexpression of the Q418A and T419A mutants in the p62 UBA construct. All mutants were expressed in C41 (DE3) *E.coli* in luria broth media. Lane 1 contains a low weight molecular weight marker used as a size standard. The total cell lysate and soluble fraction of uninduced Q418A are present in lanes 2 and 3 respectively. The total cell lysate and soluble fraction of induced Q418A are present in lanes 4 and 5 respectively. Lanes 6-9 show the same samples but for the T419A mutant.

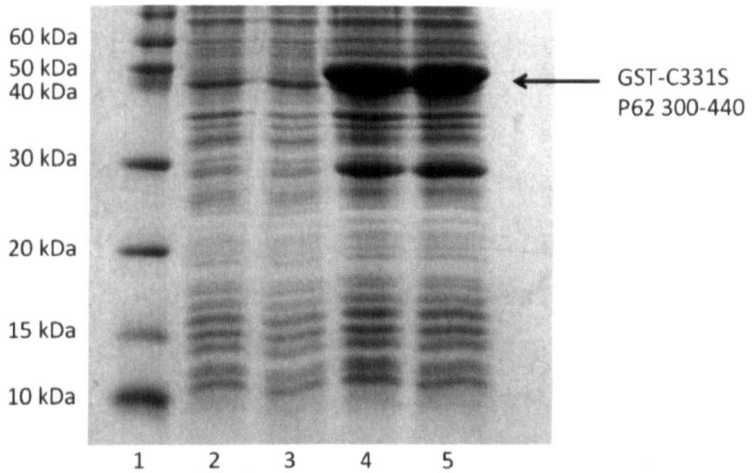


Figure 5.3 – A 20 % acrylamide SDS PAGE gel showing the overexpression of the C331S p62 300-440 construct in C41 (DE3) *E. coli* in luria broth media. A Novex prestained protein standard marker (Invitrogen) was used as a size standard (lane 1). Lanes 2 and 3 contain the total cell lysate and soluble fraction of the uninduced sample Lanes 4 and 5 contain the total cell lysate and soluble fraction of the induced sample.

5.2.4 Protein purification

All proteins were purified in three stages, a gravity flow talon column containing a glutathione Sepharose resin, gel filtration chromatography and desalting chromatography. The only difference in the protocol between cleaved proteins and GST fusions was that the proteins were either subjected to an on-column cleavage using thrombin or eluted as GST fusions from the gravity flow column. Confirmation of protein containing fractions was achieved by visualising on an SDS PAGE gel (figures 5.4 and 5.5).

Cell pellets were removed from storage at -80 °C resuspended in 6 ml Tris.HCl lysis buffer, DNaseI and a protease inhibitor cocktail. The cells were lysed to release protein into solution by sonication (30 second bursts). The cellular debris was subsequently clarified by centrifugation at 40,000 rcf for 30 minutes. The supernatant was applied an equilibrated gravity flow column and tumbled at 4 °C for 2-4 hours depending on mutant stability. The T414A, T414K, C331S p62 300-440 and the GST fusion proteins were tumbled for 2 hours, whereas all other p62 UBA mutants were tumbled for 4 hours. During this time the mutants bound to the sepharose 4B beads. Non-bound proteins were washed from the column using the same Tris.HCl lysis buffer. The column was re-equilibrated with Tris.HCl thrombin cleavage buffer. 1ml of this buffer was added to 5 units of thrombin, the mixture was applied to the gravity flow column and tumbled at 4 °C for 16 hours.

The thrombin cleaves the N-terminal GST tag leaving a GS, GSPEF or GSPNS linker attached to the protein of interest. The thrombin cleaved proteins were eluted as a 1ml solution but the column was washed with a further 1.5 ml to remove any proteins still attached to the column. GST fusion proteins were eluted from the sepharose beads by the addition Tris.HCl buffer containing reduced glutathione. The reduced glutathione out competes the GST tag for binding to the glutathione Sepharose beads. To maximise the efficiency of the reduced glutathione 6 ml Tris.HCl containing reduced glutathione was added and the column tumbled at 4 °C for 30 minutes before eluting. A further 4 ml was washed through the column to ensure proteins were completely removed from the column.

A gel filtration column was utilised to ensure any remnant proteins from the gravity flow column were removed. Since gel filtration column separate proteins

according to their size and shape any incorrectely cleaved proteins would also be removed. A Superdex 200 gel filtration column was utilised for all GST fusion proteins. A Superdex 75 column was used for the purification of the C331S p62 300-440 construct and the MAP-LC3 protein. A sephacryl S-100 column was employed to purify the various p62 UBA mutants. All cleaved proteins eluted between 170 and 230 ml. All GST fusion proteins eluted between 150 and 210 ml. A large sharp peak was recorded by monitoring the absorbance at 280 nm. The proteins were collected as 10 ml fractions, with the protein containing fractions being snap frozen in liquid N_2 prior to freeze drying for 48 hours. The MAP-LC3 protein was spin concentrated to 2.5 ml after gel filtration as it was unable to be freeze dried at this stage of the purification.

Finally a desalt column was used to decrease the NaCl content for biophysical analysis. Protein containing fractions from the gel filtration column were diluted to 2 ml and loaded onto a 5 ml HiTrap desalt column. A peak corresponding to the protein was recorded by monitoring the absorbance at 280 nm. Desalts were completed in degassed deionised water with the exception of the MAP-LC3 protein which was desalted in 25 mM ammonium acetate.

Proteins were lyophilised on a freeze drier after gel filtration and desalt steps. The MAP-LC3 purification required spin concentrating after the gel filtration as it was unable to be freeze dried at this stage of the purification. Moreover after MAP-LC3 had been desalted and subsequently freeze dried it was required to be resuspended in deionised water and freeze dried again to remove ammonium acetate that might still be bound.

Typical yields for the p62 UBA mutants were between 3-6 mg per Litre. The MAP-LC3 protein was able to obtain yields of up to 8 mg per Litre. The C331S p62 300-440 construct achieved a low yield of 3 -4 mg per Litre.

All p62 UBA mutants and MAP-LC3 were able to be purified to a single homogenous species. The GST fusion proteins and the C331S p62 300-440 construct all showed some signs of degradation. For the GST-p62 UBA, GST hHR23A UBA2 and the cleaved C331S p62 300-440 construct this degradation was minimal. However, the GST fusions of the longer p62 constructs, all showed significant degradation.

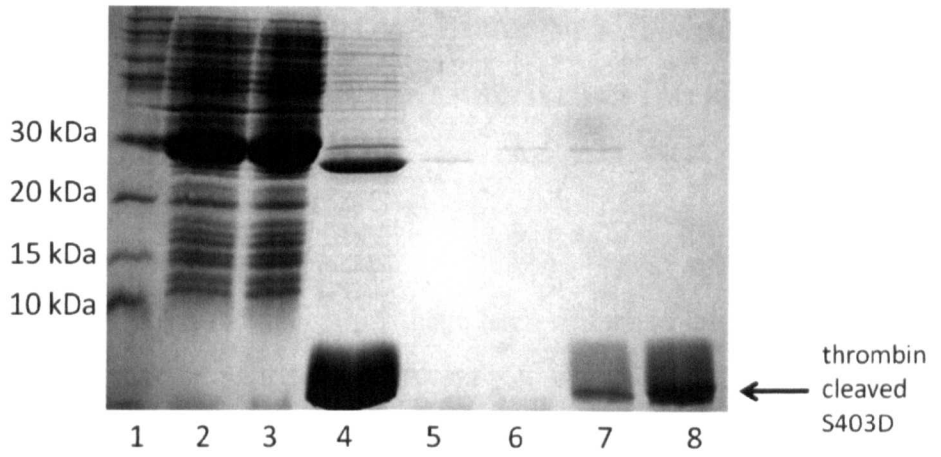


Figure 5.4 – A 20 % acrylamide SDS PAGE gel showing the purification of the S403D mutant of the p62 UBA expressed in C41 (DE3) *E.coli* in luria broth media. A Novex prestained protein standard marker (Invitrogen) was used as a size standard (lane 1). The total and soluble fractions of cell lysate are shown in lanes 2 and 3. The thrombin cleaved protein is shown in lane 4. Gel flitration fractions 16 to 19 are shown in lanes 5 to 8 respectively. Fractions 18 and 19 are the protein containing fractions.

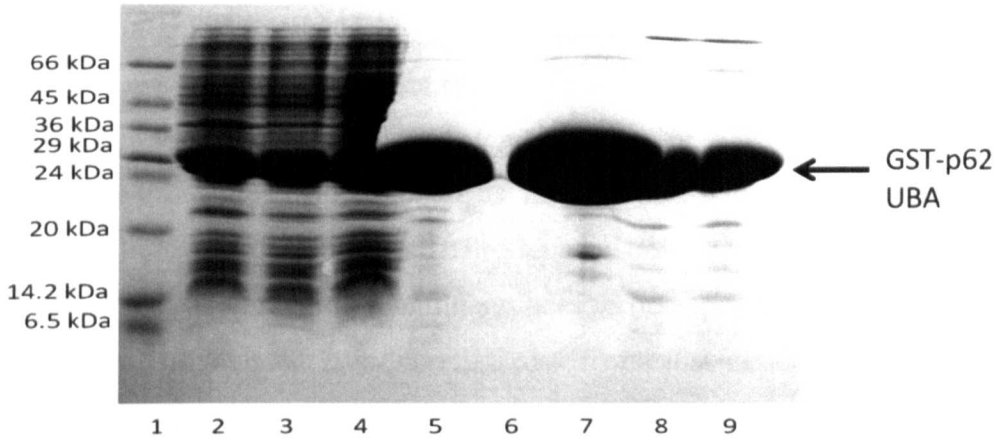


Figure 5.5 – A 20 % acrylamide gel showing the purification of GST-p62 UBA in C41 (DE3) *E.coli* in luria broth media. A low molecular weight marker was used as a size standard (lane 1). Lanes 2 and 3 correspond to the total cell lysate and soluble fraction respectively. Lane 4 is the flow through from the sepharose column. As you can see not all the protein has bound to the sepharose beads. For the GST-p62 UBA only the flow through was re-applied to the column to increase the overall yield. Lane 5 corresponds to the protein eluted from the sepharose beads. Lanes 6-9 correspond to gel filtration fractions 19-22. Fractions 20-22 are the protein containing fractions. Some slight degradation was observed but was not considered to affect overall yield.

6.0 Investigating the p62 UBA dimer by introducing mutations along helix 2 which forms the dimerisation interface.

6.1 Introduction

6.1.1 UBA dimerisation

Most of the UBA structures which have been deposited into the Protein databank form monomeric structures. Until recently it was thought that all UBA domains existed as monomers. To date structural characterization has occurred for the UBA dimers of the c-Cbl[234], Cbl-b[235], XIAP[236] and doublesex (*Drosophila melanogaster*)[237] proteins. The dimerisation interface is distinct in each of these cases, despite the c-Cbl and Cbl-b being closely related (figure 6.1). The different structures reveal different roles for dimerisation in ubiquitin binding.

Cbl proteins are a small class of E3 ubiquitin ligases which are important regulators of many receptor-tyrosine kinases. In the case of the Clb-b UBA, quite different interfaces are used for dimerisation and ubiquitin binding, with dimerisation being regulated by ubiquitin binding[235]. However, the c-Cbl UBA doesn't appear to bind to ubiquitin suggesting that the functional role of the dimer is not linked to regulating ubiquitin interactions[234]. The XIAP UBA, like the Dsk2 UBA, contains a 3_{10} helix as well as the conserved three helix bundle. The dimerisation interface is composed of residues mainly located in the 3_{10} helix and helix1, causing the ubiquitin binding surfaces to be exposed on opposite ends of the 'bowtie' structure[236]. It is therefore speculated that dimerisation confers the UBA to preferentially bind to Lys63 and linear chains over Lys48 chains. Interestingly the XIAP protein, like the p62 protein, has been linked to NF-κB activation[238]. The doublesex transcription factor rather unusually contains a UBA domain capable of dimerisation[237]. The UBA domain has been shown to be able to bind very weakly to ubiquitin using a surface that is structurally separate from the dimerisation interface. This is the first example of a UBA fold within a transcription factor.

Although currently only a few examples of UBA dimers have been reported, it is likely that many more are yet to be discovered. The function of certain UBAs may have adapted from evolutionary pressure to cause the formation of dimers, which are likely to have a downstream role in ubiquitin binding. Dimers thus far have provided evidence for both up-regulation and down-regulation of ubiquitin binding as well as being able to confer linkage specificity in polyubiquitin chain recognition. It is predicted that newly discovered UBA dimers will also have roles in regulating ubiquitin recognition and binding.

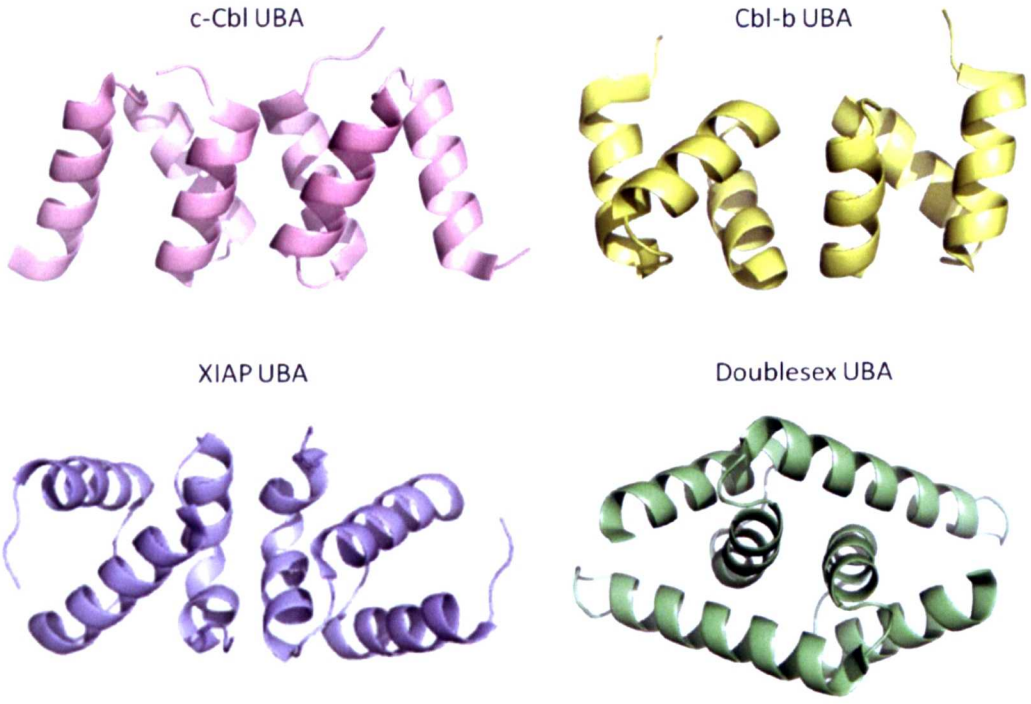


Figure 6.1 – A comparsion of the structures of different UBA dimers. The c-Cbl UBA (PDB ID 2OO9), Cbl-b UBA (PDB ID 2OOA) and doublesex UBA (PDB ID 1ZV1) were all deposited into the Protein databank as dimers. The XIAP UBA (PDB ID 2KNA) was deposited as a UBA monomer. The dimer structure was created based on the HADDOCK model described by Tse *et al*[236].

6.1.2 The p62 UBA dimer and its role in ubiquitin binding

It has emerged in the last few years that the p62 UBA domain exists as a highly stable dimer as shown by the solution NMR[141] and the 1.4 Å crystal structure[131]. The dimer is arranged with symmetry along the C2 axis. Helix 2 from each half of the dimer line up in an antiparallel fashion to form the core of the dimerisation interface. The interface itself buries an area of 667 $Å^2$ and is formed from hydrophobic residues in helices 2 and 3 (figure 6.2). The side chain from Trp 412 interacts with the side chain of Leu 416 from the other subunuit of the dimer (denoted Leu 416'). Ile 431 is found at the C terminal end of helix 3. This residue interacts with its complementary residue (Ile 431') in the other half of the dimer forming a critical hydrophobic interaction. Such hydrophobic interactions combined with the extensive size of the dimerisation interface enable the formation of a highly stable dimer. Some additional electrostatic interactions were observed in the crystal structure which would futher enhance the stability of the hydrophobic core. Met 404 bonds to Glu 432' and Arg415 bonds to Gly 410' and Gly 411'. The solution NMR structure also observes a hydrogen bond between Thr 419 and Glu 409' in an interaction that is facilitated by the presence of the di-glycine insertion in loop 1, a feature unique to p62 UBA.

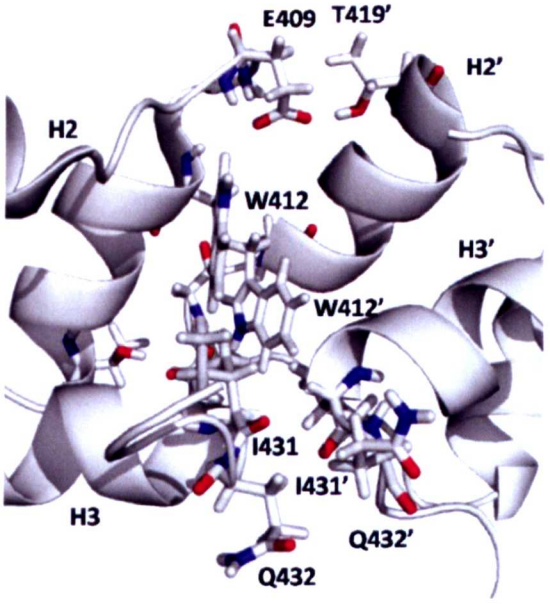


Figure 6.2 – a zoom of the p62 UBA dimerisation interface (PDB ID 2KNV). Residues which stabilize the interface are highlighted and with their side chains shown as sticks

The p62 UBA dimer has an apparent K_d for the monomer-dimer equilibrium (K_{dim}) of 4.1 μM, indicating a very strong preference for the dimeric species at physiological concentrations. The dimer dissociates in the presence of ubiquitin to an unstable monomeric form which subsequently binds to ubiquitin to form a stable complex (figure 6.3). The dissociation and binding events are mutually exclusive due to the partial overlap of the interfaces. It is this overlap which allows the dimer to tightly regulate the affinity for ubiquitin. The unstable monomer has been shown to be the biologically active form because only this form is capable of binding to ubiquitin. The p62 UBA monomer binds to the canonical binding patch centred on Ile44 and Val 70 on ubiquitin and involves a conserved MGF motif in loop 1 and residues in helix 3 of the UBA. The interaction with ubiquitin utilises the hydrophobic patches on both proteins.

Some structural differences have been observed between the dimer and monomer, the most notable being an extension of helix 3 in the monomer. In the dimeric form helix 3 is defined by residues Gly 425 to Ile 431; however, in the monomeric form helix 3 extends by half a turn to terminate at Tyr 433. This extra half turn in the dimer prevents association of the two subunits. This helical extension represents another level of control by the p62 in regulating ubiquitin binding.

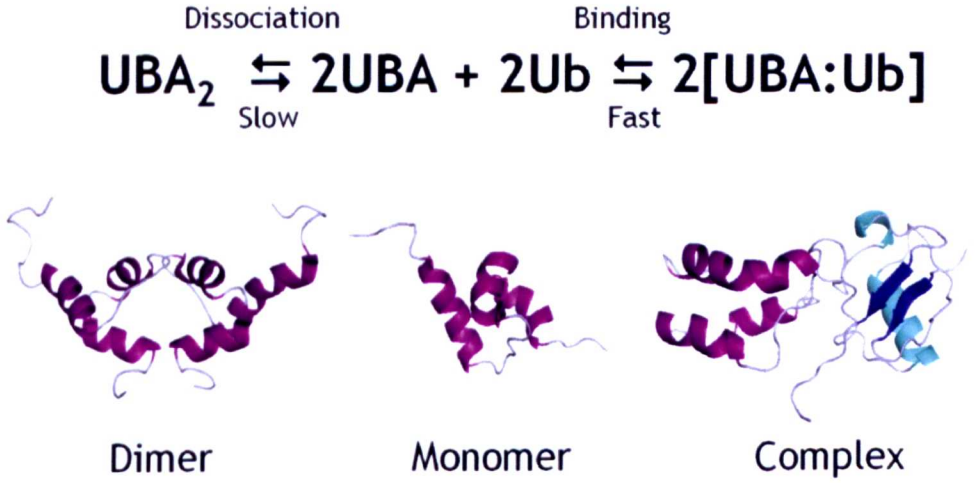


Figure 6.3 – The mechanism of binding by the p62 UBA. The highly stable dimer (PDB ID 2KNV) exists in solution but in the presence of ubiquitin the dimer dissociates to the free monomeric form (PDB ID 1Q02) which is then able to bind forming a complex. The complex is a representation made from the PBDs of ubiquitin and the bound monomer (PDB ID's 1UBQ and 2JY8).

6.1.3 Summary and aims

With the discovery of the p62 UBA dimer, we aim to probe the dimerisation interface. By introducing mutations along helix 2, which forms the dimerisation interface (figure 6.4), it is postulated that weaker dimers and monomeric UBAs will be produced. A study conducted in 2010 highlights Lys as being the least favourable interface residue[239]. Charged residues, such as Lys, contribute greatly to the accessible surface area. The bulky lysine side chain is therefore able to cause steric clashes with residues on the other half of the dimer, disrupting the interface. The mutants T414K, R415K, Q418K and T419K were therefore generated (table 6.1). These residues are surface exposed in the wild type dimer, allowing mutation of these residues to affect the electrostatic interactions formed at the dimerisation interface. The corresponding Ala mutants, namely T414A, R415A, Q418A and T419A, had been previously designed when searching for a secondary ubiquitin binding site on the UBA before the discovery of the dimer. By mutating to Ala any side chains capable of forming stabilising interactions at the interface are removed, which in theory can disrupt dimer formation. These four mutants were therefore also generated (table 6.1). Of these eight mutations the T419K mutant was predicted to be the most destabilising. The T419 residue forms a hydrogen bond with the E409 residue in the opposite half of the dimer (denoted E409') stabilising the interaction at the interface (figure 6.4).

Table 6.1 – A list of all the p62 UBA helix 2 mutants. The mutated residue is shown for each mutant in red.

Mutant	Helix 2 sequence (residues 413 – 420)
T414A	L A R L L Q T K
R415A	L T A L L Q T K
Q418A	L T R L L A T K
T419A	L T R L L Q A K
T414K	L K R L L Q T K
R415K	L T K L L Q T K
Q418K	L T R L L K T K
T419K	L T R L L Q K K

Perhaps the most interesting feature of the p62 UBA is that in order to bind to ubiquitin the dimer must first dissociate. It is the regulation of binding by dimerisation that is the focus of this part of the project. Knowledge gained on the monomer-dimer equilibrium and its role in modulating ubiquitin binding may

provide a key link to the impaired ubiquitin binding experienced by patients with Paget's disease. With so many PDB mutations now known, combined with the fact that most of the mutations are not confined to the ubiquitin binding patch, highlights the need to better understand the molecular mechanisms which underlie PDB. It is clear that structural alterations in the ubiquitin binding patch are not solely responsible for defective ubiquitin binding and that other factors, including dimerisation, must also play a role.

By generating mutants designed to disrupt dimer formation and characterising them using structural and biophysical approaches, a greater understanding of the role of the dimer in ubiquitin recognition can be gained. The effects of dimerisation can subsequently be explored in the PDB mutants. If these mutants are able to prevent dimer formation and exist as monomers, the biologically active form could also be characterised. Little is known about the wild type monomer due to its low population at equilibrium and high instability.

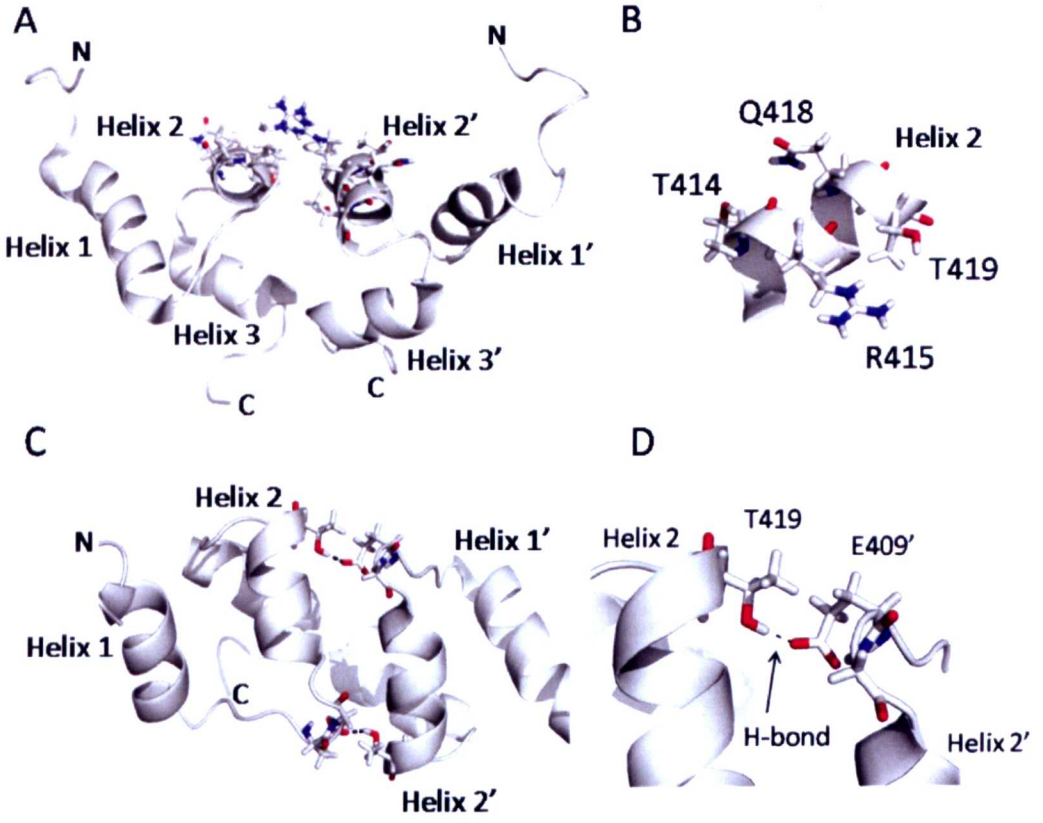


Figure 6.4 – The p62 UBA dimerisation interface (PDB ID 2KNV) A) the side chains of the surface exposed residues which were mutated in this study are shown as sticks. B) Helix 2 rotated by 90° horizontally and zoomed. C) The dimerisation interface rotated by 90° vertically to highlight the T419, E409 and G410 residues at the ends of helix 2. D) A zoom of the hydrogen bond which is formed between T419 and E409'.

6.2 Results

6.2.1 Biophysical studies on the monomer-dimer equilibrium

6.2.1.1 Far UV-CD dilution experiments

In order to probe the effects of mutation on the thermal stability of the helix 2 mutants, concentration dependent thermal unfolding studies were performed using Far UV-CD spectroscopy. The far UV-CD spectrum of a protein is highly sensitive to perturbations in secondary structure. Small changes in secondary structure can lead to large changes in the CD signal. Although Far UV-CD does not provide atomic level detail an average measure of the change in the overall structure can be provided. Far UV-CD is therefore an ideal technique to study the thermodynamic properties or folding mechanisms of proteins.

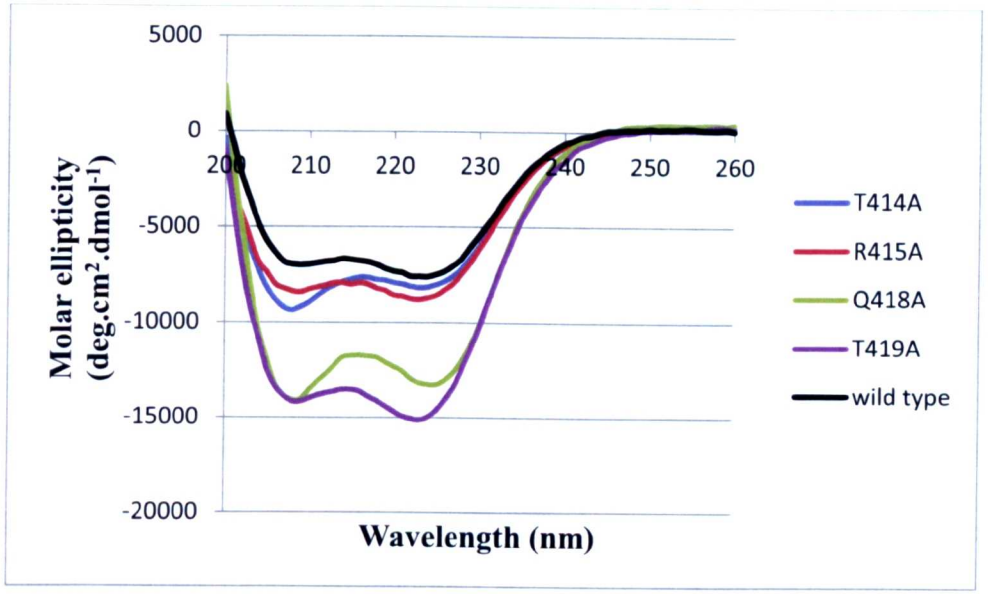
Far UV-CD scans were recorded between 200-260 nm at temperatures of 298K prior to thermal melting. These scans checked for helical content to ensure the mutants were folded into the native state prior to thermal denaturation. The ellipticity at 222 nm was subsequently recorded as a function of temperature at 0.5 K intervals between 278.15 K and 368.15 K. The apparent midpoint of the unfolding transition (T_m) was calculated from these experiments using the first derivative method suggested by Greenfield et al[185]. Initial concentrations (170 μM – 200 μM) were serially diluted by a factor of two over eight different concentrations. At higher concentrations the dimer is favoured whereas at lower concentrations a shift in the monomer-dimer equilibrium occurs to favour the monomer. A shift in the T_m value in a concentration dependent manner therefore represents a shift in the equilibrium. Thus multiple species in equilibrium can be detected using these thermal unfolding profiles. Experiments on the wild type p62 UBA have been previously completed by the Searle group allowing the thermal stabilities (T_m) of the mutants to be compared to the values observed for the wild type.

A far UV-CD spectrum which corrects for wavelength and concentration was visualised by plotting Molar ellipticity against wavelength. The Far UV-CD spectrum produced for each mutant across the concentration range exhibited strong negative bands at 222 nm and 208 nm, indicating a large helical content. The band at 222 nm is observed due to the strong hydrogen bonding environment of α helical conformations, whereas the intensity at 208 nm is associated with the length of α helices[240]. The far UV-CD spectra indicate that the compact three helical bundle is maintained for all p62 UBA dimer mutants (figure 6.5). The negative bands at 208 and 222 nm are observed at each concentration throughout the course of the dilution highlighting that both species, the dimer and the monomer, stayed largely helical.

Despite this, some differences were observed by some of the p62 UBA mutants compared to the wild type p62 UBA. The T414A and R415A p62 UBA mutants produced curves which were highly similar to wild type suggesting that these mutations have little effect on the overall structure of UBA domain. All other p62 UBA mutants produced noticeably stronger bands at 208 and 222 nm, indicating that these mutations have a greater effect on the helical content of the domain. In general the intensities at 208 and 222 nm are used to estimate helical content[240]. The T414K mutant exhibits the strongest negative bands at 222 and 208 nm, suggesting that this mutation has a more drastic effect compared to the other mutations.

Since the recorded far UV-CD spectrum for all of the p62 UBA mutants at each concentration showed the characteristic negative bands at 208 and 222 nm indicative of a solely helical protein, it is highly likely that the mutations have caused local changes to structure only. A rearrangement of the protein backbone is unlikely given the consistency of the spectra with helical conformations across the concentration range. The CD signal is extremely sensitive to small structural perturbations, making it an ideal technique to monitor secondary structure. However, experiments using techniques which have atomic resolution such as X-ray crystallography or NMR would be required to probe any detailed local changes to structure.

A



B

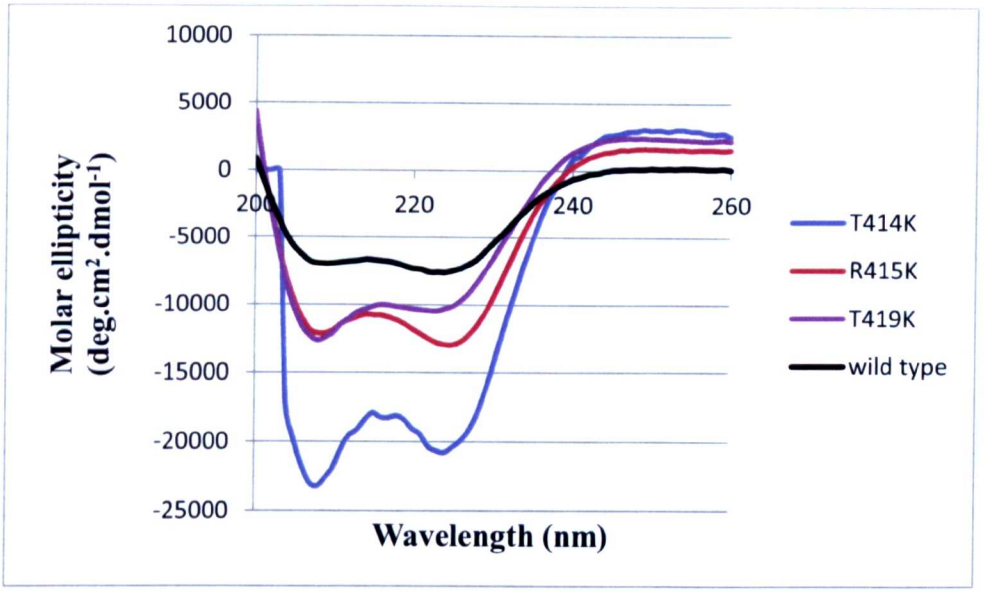


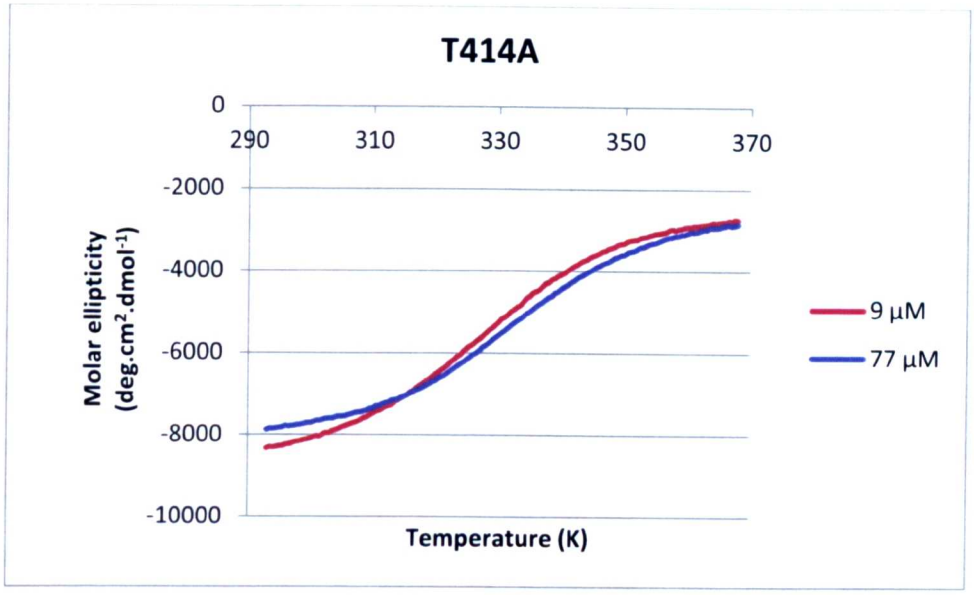
Figure 6.5 – The far UV-CD spectrum of the p62 UBA helix 2 mutants. A) p62 UBA Ala mutants (T414A, R415A, Q418A and T419K) .B) p62 UBA Lys mutants (T414K, R415K and T419K). The strong negative bands at 208 and 222 nm indicate an α helical conformation. The wild type data is shown on both spectra as a comparison. The data was recorded in 10 mM potassium phosphate, pH 7 and 298K.

A fully folded protein was required to ensure an accurate T_m value was calculated from the thermal unfolding studies. Once the Far UV-CD scans had been completed to check for helical content, a thermal melt ensued. The thermal stabilities of the mutants were measured as a function of temperature in the unfolding profiles (figure 6.6). Each mutant was serially diluted by a factor of two over eight different concentrations. During the dilution a noticeable change in thermal stability occurs at lower concentrations (< 30 μM). The change represents the presence of a second species, in this case the appearance of the monomer. The T_m values for the monomer and dimer of each mutant are listed in table 6.2 and are shown graphically in figure 6.7.

The thermal unfolding data suggests that the majority of mutations are able to destabilise both the dimer and monomer species. This is shown by the decreased values of T_m when compared to wild type protein. The only exception is the T419A mutant which showed an increase in the thermal stability for both the dimer and the monomer. The wild type p62 UBA showed a sharp decrease in thermal stability (a change of 6 K) over the concentration range[141], representing the transition from dimer to predominantly monomer. The R415A, Q418A and T419A mutants also showed a strong decrease in thermal stability over the concentration range, with changes between 5 K and 8 K observed for these mutants.

However, the R415K and T419K mutants show a much weaker change in thermal stability as they only change by 2 K and 3 K respectively. The T419K monomer showed a T_m value that is only 2 K lower than the wild type monomer. However, the difference in the T_m values observed for the dimers was much greater at 5 K. This would suggest that the T419K mutation has had a much greater effect on the thermal stability of the dimer, consistent with the breaking of the hydrogen bond formed by the wild type p62 UBA dimer. The results for the R415K mutant are similar to the T419K mutant, except both species show a greater destabilization than the T419K. This is unusual for a like-for-like mutation such as R415K.

A



B

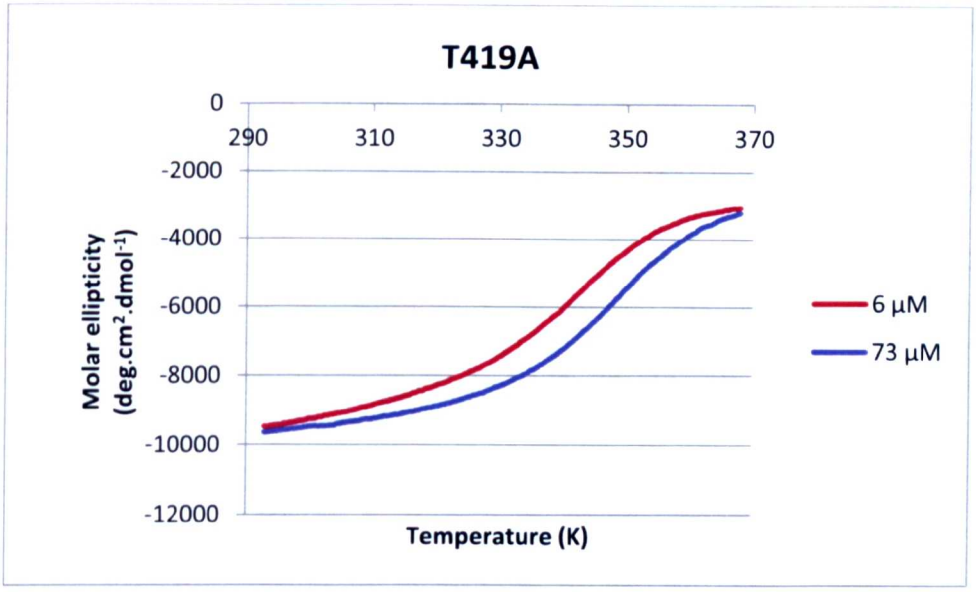


Figure 6.6 - Far UV-CD thermal unfolding curves recorded at 222 nm for A) the T414A mutant at 9 μM and 77 μM and B) the T419A mutant at 6 μM and 73 μM. The curve for the lower protein concentration in each case has been scaled to show approximately the same change in Molar ellipticity. The T419A mutant shows unfolding curves which highlight the difference in the midpoint of unfolding at different concentrations; whereas, the T414A has a similar midpoint of unfolding at high and low concentrations.

Interestingly, the two mutants at position 414, T414A and T414K both show no real evidence of a change in thermal stability. This means that the T_m value does not noticeably change across the concentration range. The T_m values calculated for the monomer and dimer of these mutants are substantially reduced compared to the wild type, suggesting considerable thermal destabilization in these mutants. Thr414 is therefore likely to have a role in maintaining the stability of the both species. The lack on an observable change in T_m suggests the presence of a single species rather than an equilibrium between two species. The more likely scenario is that the thermal stabilities of the monomeric and dimeric species of these mutants are therefore very similar to each other making them indistinguishable using this technique. If this is the case, the T414A and T414K dimers, have been more dramatically affected than the monomer, as the wild type dimer is more thermally stable than the monomer.

Table 6.2– The apparent T_m values for the dimer and monomer using Far UV-CD thermal unfolding curves. The T_m values were calculated using first derivatives method outlined by Greenfield *et al*[185].

Mutant p62 UBA	T_m dimer (K)	T_m monomer (K)	ΔT_m (K)
Wild type	345	339	6
T414A	331	331	0
R415A	343	335	8
Q418A	338	333	5
T419A	350	342	8
T414K	324	324	0
R415K	338	336	2
T419K	340	337	3

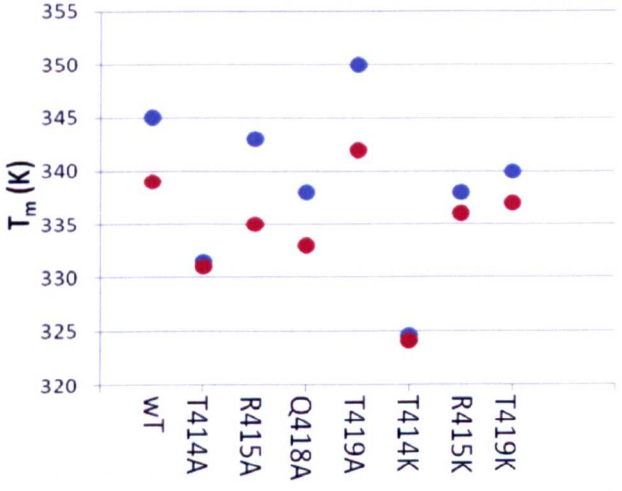


Figure 6.7 – Concentration dependence of the T_m values recorded for the helix 2 mutants, showing the upper (T_m dimer) and lower limits (T_m monomer) of thermal stability over the concentration range of 1 -200 μM in blue and red respectively.

6.2.1.2 ITC dilution experiments

The data produced from the thermal unfolding CD studies indicated the presence of a single species in equilibrium for the T414A and T414K mutants. In order to check the monomer-dimer equilibrium exists for all of the helix 2 mutants, dilution studies using ITC were conducted. When substances dissociate heat is absorbed. That heat can be measured by the ITC calorimeter to produce a dissociation curve. Dilution ITC is therefore an suitable technique for probing dimer dissociations. If a significant endothermic heat response is not observed by the T414A and T414K mutants it is likely that these mutants are monomeric. Proteins which are monomeric, such as the hHR23A UBA2, produce only a small exothermic heat response after each injection[241], generating a trace which is indistinguishable from injecting water into water.

Concentrated samples between 240 μM and 260 μM were prepared in 50 mM potassium phosphate, 50 mM NaCl, pH 7 and sequentially injected to a cell containing degassed buffer. Upon entering the low concentration of the cell the dimers dissociate. As the concentration of the cell increases the endothermic heat spikes decrease in intensity until dissociation in the cell is not longer significant. The resultant dimer dissociation curve was fitted to a dimer dissociation model to produce accurate K_{dim} and ΔH values (figure 6.8). From these values ΔG and ΔS values can be calculated to generate a full thermodynamic profile.

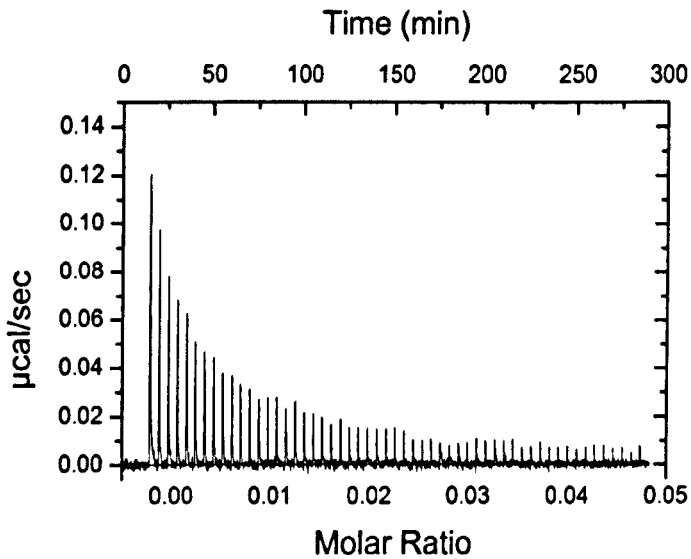


Figure 6.8 – The raw ITC dissociation curve for the R415A p62 UBA mutant at 240 μM. This dissociation curve is representative for all the helix 2 p62 UBA mutants. Experiments were performed in 50 mM potassium phosphate, 50 mM NaCl, pH 7.

All data, except those recorded for the T419K mutant, fitted well to a dimer dissociation model. The T419K mutant produced slightly confusing results despite repeating the experiment across a range of concentrations (1 mM, 500 μM and 250 μM). The initial 2-3 injections showed a slight increase in the intensity of the endothermic heat spikes before the normal dissociation curve resumed. This prevented the data from being accurately fitted. A poor fit was able to be produced from an experiment conducted on a 250 μM sample by removing the first data point.

The K_{dim} values for the R415A, Q418A, T419A and R415K mutants were in excellent agreement with the K_{dim} observed for wild type p62 UBA. The wild type p62 UBA was previously shown to have a K_{dim} of 4.1 ± 0.6 μM[141]. Significantly weaker K_{dim} values were recorded for the T414A, T414K and T419K mutants. The value for the T419K mutant is likely to be less accurate due to the poor fit of the dissociation curve; however, if this mutant was wild type-like then it would have produced better data.

The thermodynamic profile for each of the helix 2 mutants is shown in table 6.3. The thermodynamic profile for the dissociation process revealed positive values for ΔH and ΔG and negative values for ΔS. This highlights that dimer dissociation is not spontaneous and that reverse process, the association of two monomers, is the spontaneous process. This suggests that the monomeric form is less stable than the dimer, consistent with previous findings for wild type p62 UBA[141]. Although the monomer is unstable, the negative value for ΔS indicates a favourable change in entropy occurs during dimer dissociation. The larger value for ΔH compared to ΔS suggests that the reaction is enthalpy driven. When you compare the thermodynamic profiles of the p62 UBA dimer mutants, the values for ΔH are quite similar suggesting a similar amount of energy is required for dissociation. However, the ΔS values are significantly lower for the T414A, T414K and T419K p62 UBA mutants indicating changes in the distribution of thermal energy of these mutants indicative of a substantial change in the populations of monomer and dimer.

Table 6.3 – The enthalpy change and dissociation constants for the helix 2 mutants calculated from dilution ITC experiments and fitted using a dimer dissociation model.

Mutant	K_{dim} (±)/ μM	ΔH (±)/ J/mol	ΔS (±)/ J/K/mol	ΔG_{298} (±)/ KJ/mol
Wild type	4.1 ± (0.6)	60.7 (±1.0)	-102.93 (±1.3)	30.7 (±0.3)
T414A	141.0(± 1.0)	68.9 (±1.0)	-73.48 (0.06)	22.0 (±0.02)
R415A	3.8 (±0.4)	64.9 (±2.2)	-103.59 (±0.9)	30.9 (±0.3)
Q418A	3.7 (±0.4)	79.3 (±2.7)	-103.71 (±0.9)	31.0 (±0.3)
T419A	4.5 (± 0.5)	45.2 (±1.5)	-102.22 (±0.9)	30.5 (±0.3)
T414K	146.0 (± 1.5)	67.5 (±1.5)	-73.20 (±0.09)	21.9 (±0.03)
R415K	3.7 (± 0.5)	65.1 (±2.6)	-103.81 (±1.04)	31.0 (±0.3)
T419K	165.0 (±30.0)	38.1 (±1.6)	-72.28 (±1.39)	21.6 (±0.4)

Overall the dilution ITC data is in good agreement with the thermal stabilities produced from the far UV-CD unfolding studies. The T414A, T414K and T419K mutants all show larger values of K_{dim} and reduced thermal stabilities suggesting that weaker dimers are formed which are not as stable as the wild type protein. This was expected for the T419K mutant due to the breaking of the hydrogen bond it normally forms with E409'. Residue T414 is located at the N-terminal end of helix 2, mutation of this residue could cause larger structural perturbations than any of the other mutations, such as helical unwinding, although NMR experiments would need to be conducted to confirm this.

6.2.1.3 ESI-MS dilution experiments

ESI-MS is a type of mass spectrometry which has been developed to observe macromolecules and macromolecular complexes *in vacuo*. Molecules are transferred to the gas phase without fragmenting using a fine aerosol known as electrospray. Samples are dissolved in a volatile solvent, such as ammonium acetate, and pushed through a small capillary where a high voltage is applied causing the formation of ionized droplets. The droplets reduce in size due to solvent evaporation and eventually are separated by a mass analyser according values of mass to charge ratio (m/z) and detected by a detector. A series of different charge states at different relative abundances are observed and a spectrum representing a "snap shot" in time is visualized. From the spectrum produced a variety of information can be deduced, including information on non covalent protein-protein interactions and stoichiometry of molecular assemblies as well as simple confirmation of molecular weights.

Initially it was important to check the molecular weight of the mutants to confirm the successful insertion of the mutation to the purified proteins. Samples were prepared by dissolving lyophilised protein in 25 mM ammonium acetate, pH 7. The molecular mass of each of the mutants was confirmed using native ESI-MS on a 5 μM sample. This technique allowed more accurate confirmation than visualisation by SDS PAGE alone. The observed masses were in good agreement with the predicted masses for all seven helix 2 mutants (table 6.4).

Table 6.4 – The molecular weights observed by ESI-MS compared to the predicted molecular weights. Both weights correspond to the UBA monomer

Mutant	Predicted mass of monomer	Mass observed by ESI-MS
T414A	5709.4	5714.2
R415A	5654.3	5659.0
Q418A	5682.3	5687.2
T419A	5709.4	5714.2
T414K	5766.5	5771.3
R415K	5711.4	5716.2
T419K	5766.5	5771.4

For all mutants, the observed spectrum was very similar. Ions corresponding to the +2 to +6 charge state were observed for the monomer. The observation of a dimeric species was apparent from the +5 charge state in which a corresponding peak for the monomer would exist as an impossible half charge state. The

intensity of the +5 ion for the dimer is low and no other ions corresponding to odd numbered charge states of the dimer were observed, highlighting the low abundance of the dimer in the gas phase. During ionization, the dimeric species is unfavoured as the UBA dimer is stabilized by hydrophobic contacts which in the absence of a polar solvent are considered to be much weaker and can be very difficult to study in the gas phase. The visualization the dimer, albeit low, is likely to be the result of the stabilizing electrostatic interactions which are also formed at the dimerisation interface. The charge state distributions and the dominant monomeric species observed for the mutants is consistent with the data previously recorded for the wild type p62 UBA. Also no higher order oligomers were detected at this concentration, again consistent with data previously collected for the wild type protein. The charge states are low with a low distribution indicating that both species are likely to be folded[242]. Although the monomeric species was the visually dominant species, accurate quantification of the two species was precluded by overlap of peaks where the mass and the charge were both doubled giving the same m/z value.

In the case of these experiments the monomer-dimer equilibrium was the focus of study. In order to probe this equilibrium the helix 2 dimer mutants were subject to dilution ESI-MS experiments (figure 6.9). Each mutant was monitored in the concentration range of 8 μM to 1 μM. Ions were observed for both the monomeric and dimeric species in distributions as outlined above. Again no higher order oligomers were observed. The overlap of most peaks in the spectrum prevented the accurate determination of the populations of both species. Despite this, changes in the populations of the monomer and dimer can be observed as the intensities of the mass ions change during the dilution. A value for the average charge state can therefore be calculated from the observed ions in each spectrum. In this case, the average charge state of the monomer was calculated as this was the predominant species. The average charge state is believed to be proportional to the solvent exposed area during ionization[221]. The monomer and the dimer have different solvent exposed surface areas, meaning that they would have differing average charge states. As the concentration decreases, the equilibrium would be expected to favour the monomer with concentration dependent changes in the average charge state reflecting an increase in the populations of monomer.

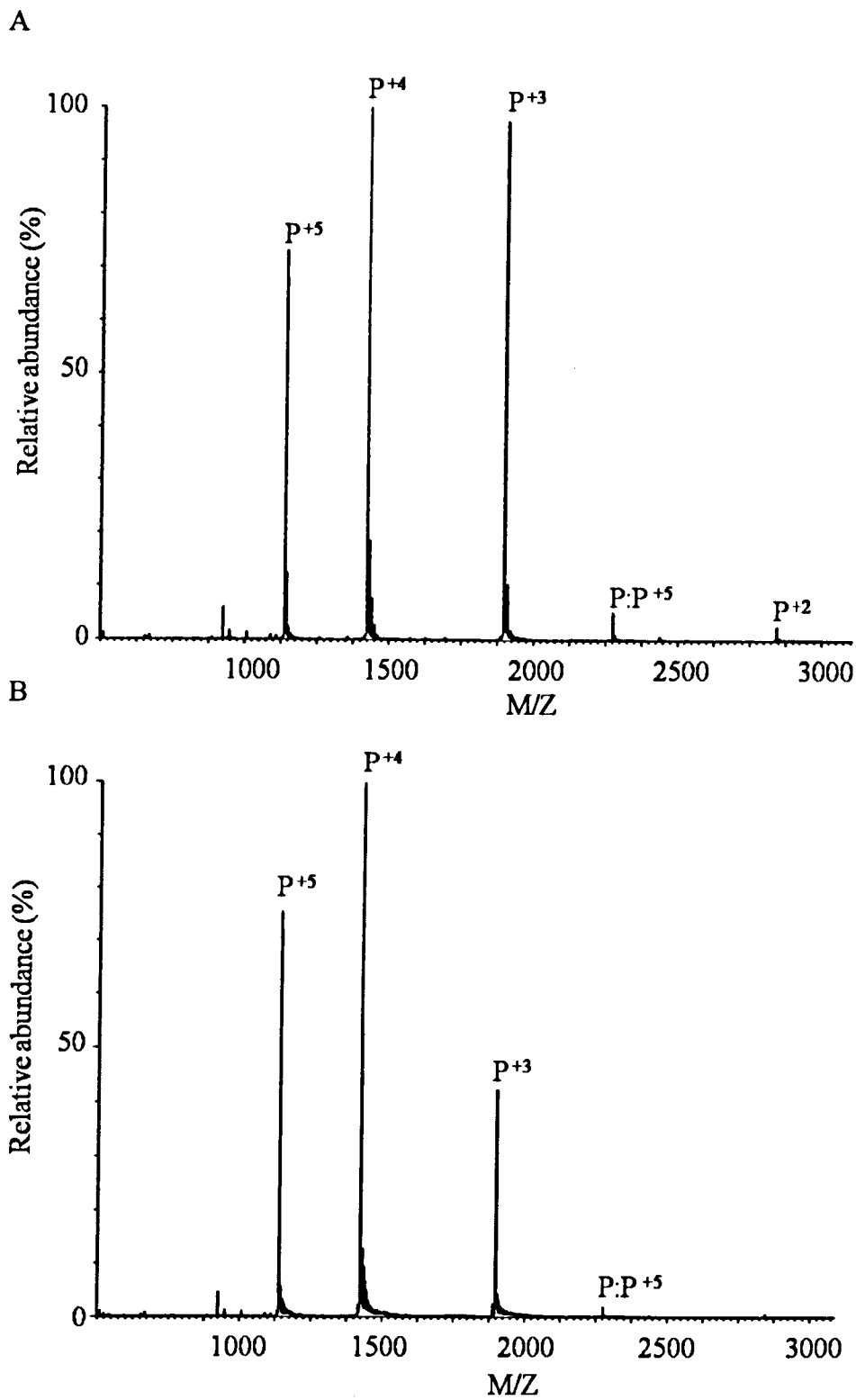
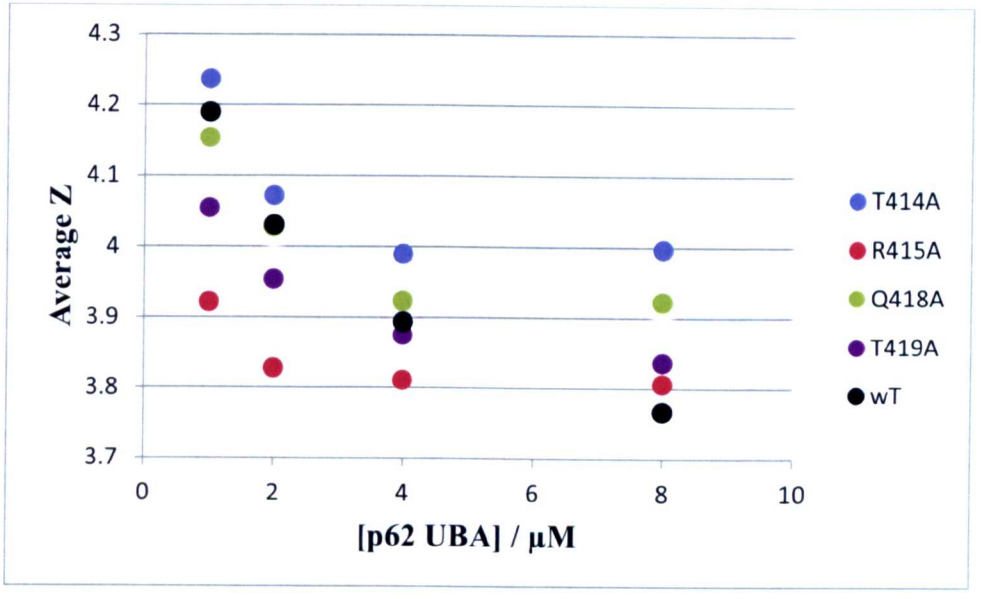


Figure 6.9 – The native ESI-MS of the Q418A p62 UBA mutant at A) 8μM and B) 1μM. Spectra were recorded in 25 mM ammonium acetate, pH 7. The labels correspond to the dominant monomeric species (denoted P) except for the +5 ion for the dimer (denoted P:P) which doesn't overlap with a monomer peak. The changes in the monomer-dimer equilibrium are easily visualized in these two spectra by the changes in intensities by the observed ions.

The average charge state was plotted against the concentration of the sample and the values compared the data to that previously reported for the wild type p62 UBA (figure 6.10). The average charge state for the wild type p62 UBA showed a clear concentration dependence. As the concentration decreases, the average charge state increases, reflecting that the solvent exposed surface area increases. This is consistent with a shift in the equilibrium to favour the monomer. The mutants also showed an increase in the average charge state as the concentration decreases, suggesting that the equilibrium is shifting to favour the monomer at lower concentrations as it does for the wild type protein. Despite the correlation between the mutants and the wild type, the magnitude of change in the observed concentration range is not as large as it was for wild type p62 UBA. This highlights a weaker concentration dependence. The only mutant with a similar magnitude of concentration dependence is the T419K mutant.

The Q418A mutant show average charge states similar to those observed for wild type protein suggesting that this mutation has had very little effect on the monomer-dimer equilibrium. Interestingly the R415A, T419A and R415K mutants show decreased values for the average charge state when compared to wild type p62 UBA, suggesting that these mutants have a decreased solvent exposed surface area. The T414A, T414K and T419K mutants all show significantly larger values for the average charge state when compared to the wild type protein, highlighting an increase in the solvent exposed surface area for these mutants. This data therefore suggests these mutants have significantly reduced populations of dimer compared to wild type p62 UBA. Alternatively partial unfolding could responsible for the increase in the solvent exposed surface area.

A



B

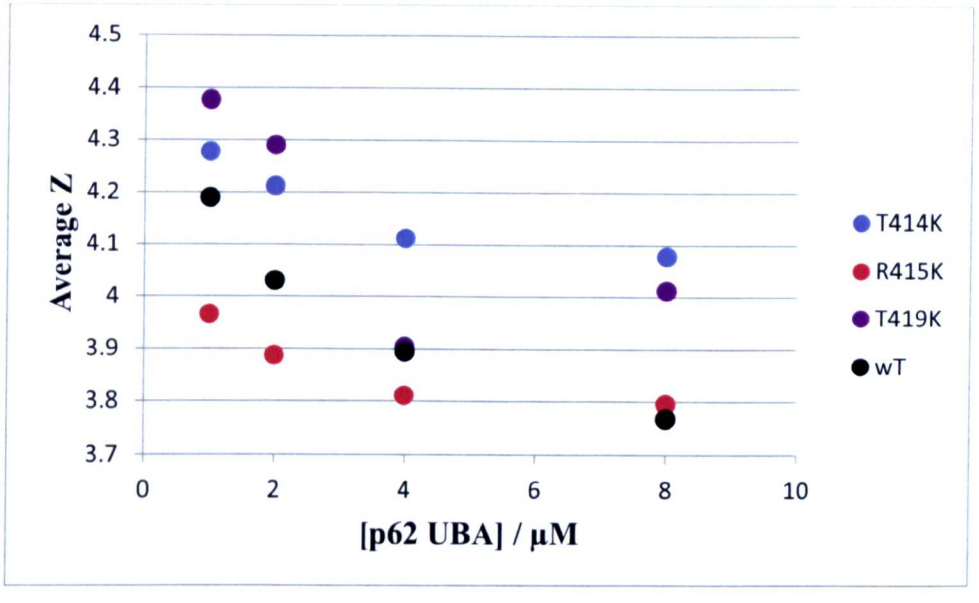


Figure 6.10 – The average charge state was calculated from the intensity of the ions recorded in the ESI-MS spectrum. Spectra were recorded for mutant p62 UBA samples at 8 µM, 4 µM, 2 µM and 1 µM. The diagram shows the average charge state (Z) plotted against concentration of p62 UBA: A) the p62 UBA Ala mutants and B) the p62 UBA Lys mutants. The wild type (wT) data is also shown as a comparison. The data shows that as the concentration decreases the average charge state increases. This change represents the change in populations of monomer and dimer at each concentration. Also the T414A, T414K and T419K mutants have a higher average charge state than wild type p62 UBA, indicating that these mutants have an increased solvent exposed surface area.

The dimer was detected in much lower abundance than expected. The low abundance of dimer was not evident from the far UV-CD and ITC dilution studies for the mutants or any of the previous work conducted on the wild type p62 UBA. It was therefore decided to repeat the native ESI-MS dilutions in the presence of a small reservoir of acetonitrile (5 ml). The presence of a neutral solvent, such as acetonitrile, in the electrospray source has the effect of lowering the charge states and decreasing the intra-molecular and intermolecular repulsion between two positively charged species generated by ionization[243]. In the case of the p62 UBA the two monomeric units appear to dissociate due to the effect of Coulomb repulsion causing low detection of the dimer which is not observed outside of the gas phase. By adding acetonitrile to the ionization source a more native-like distribution of species, as present in solution, should be observed.

The same samples prepared for the native ESI-MS experiments were used to repeat the dilution experiments but in the presence of acetonitrile (figure 6.11). The spectra recorded for the mutants were very similar. A narrower charge state distribution was observed compared to the native ESI-MS spectra, whereby ions corresponding to the monomer were seen for the +2 to +4 charge states. The +3 charge state was the most populated ion in all cases. The +5 charge state of the dimer was also present and in significantly higher abundance than in the native ESI-MS spectra. However, the data still indicated that the monomer was the dominant species in the gas phase, despite the acetonitrile increasing the amount of dimer present. The dimer is largely stabilized by hydrophobic interactions at the dimer interface. Hydrophobic interactions have been shown to be significantly weakened in the gas phase and although the acetonitrile has the ability to reduce the effects of columbic repulsion it is unable to prevent hydrophobic interactions from dissociating in the gas phase.

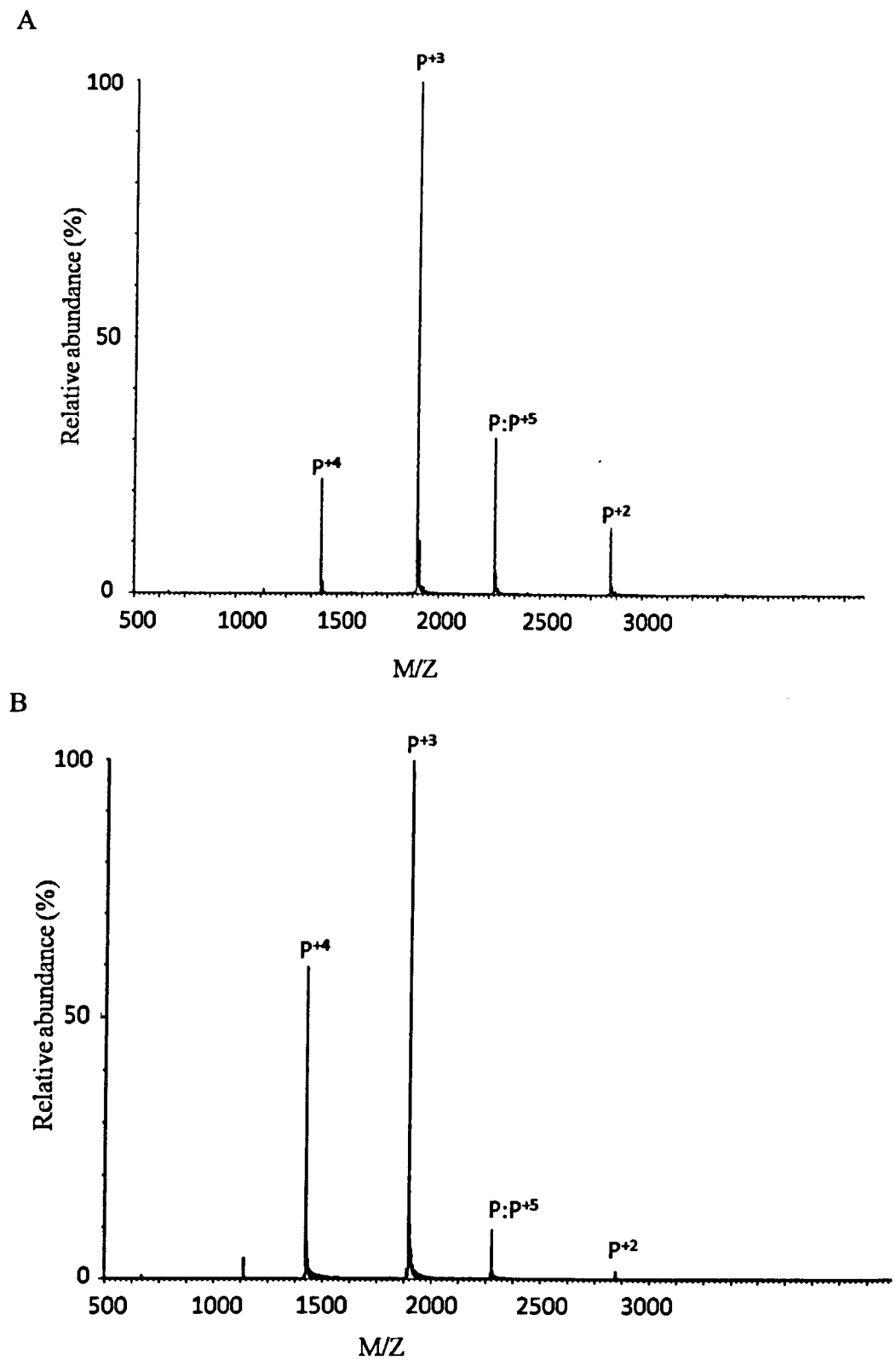


Figure 6.11 – The native ESI-MS of the Q418A mutant at A) 8μM and B) 1μM. Spectra were recorded in 25 mM ammonium acetate, pH 7. A small reservoir of neat acetonitrile was placed in the electrospray source. The labels correspond to the dominant monomeric species (denoted P) except for the +5 ion for the dimer (denoted P:P) which doesn't overlap with a monomer peak. The changes in the monomer-dimer equilibrium are easily visualized in these two spectra by the changes in intensities by the observed ions.

6.2.2.4 NMR dilution experiments

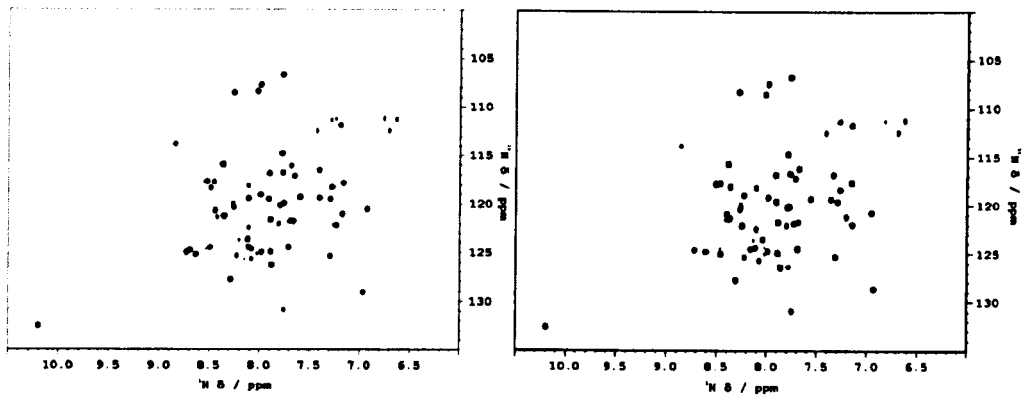
In order to probe the effects of mutations on the structure of the protein experiments using NMR spectroscopy ensued. The advantage of using NMR over the biophysical techniques is that data is produced at an atomic level. Residues with chemical shift deviations compared to the wild type can readily be identified as a full assignment of the p62 UBA dimer and monomer have already been reported[141]. Only the weak dimer mutants, namely the T414A, T414K and T419K mutants, were investigated by NMR as these are the one which the biophysical data indicated formed weaker dimers. All NMR experiments were conducted in 25 mM potassium phosphate, 25 mM NaCl, pH 7. All experiments were performed at 298 K. The buffer and temperature conditions had been optimised previously by Dr Thomas Garner[241].

Heteronuclear ^{1}H-^{15}N-HSQC experiments were completed on the weak dimer mutants at a range of concentrations between 500 μM and 10 μM (figures 6.12 and 6.13). The T414A and T414K mutants were sensitive and low yielding mutants therefore only a high concentration (500 μM) and a low concentration (10 μM) sample were prepared in the interest of saving time and minimizing protein degradation. Wild type p62 UBA had been previously shown to be fully dimeric at higher concentrations. Data was recorded on the wild type protein and was in excellent agreement with data previously recorded by our collaborator Dr M. Pandya at the University of Sheffield. The populations of monomer and dimer were shown to vary across the concentration range. At 6 μM the wild type protein shows roughly equal populations of monomer and dimer exist at equilibrium. However when diluted to 1.5 μM the equilibrium shifts towards the monomer with the monomer now being the dominant species (approximately 80 % monomer).

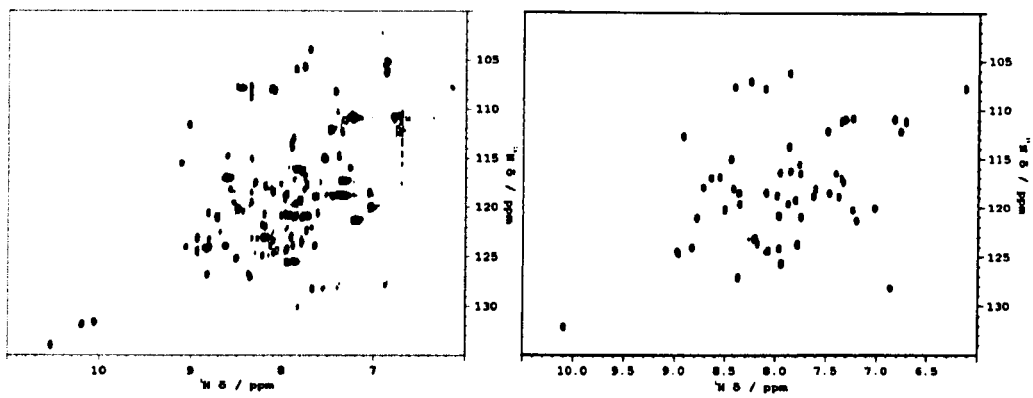
The ^{15}N-T419K mutant was diluted to 12.5 μM across a range of concentrations. At higher concentration (> 50 μM) peaks corresponding to three species at equilibrium were visible. The intensity of the peaks showed concentration dependence, which is consistent with a monomer-dimer equilibrium. The three species were easily identified from the three clearly distinct Trp side chain peaks. The third species showed a chemical shift distribution similar to wild type and it maintained the C2 symmetry as only a single set of peaks were observed. The

presence of an additional species had been previously observed by the E409K and G410K p62 UBA mutants[141]. The three species were identified as the native monomer and dimer, with the third species corresponding to a second uncharacterised dimer. As the concentration decreases, the intensity of the peaks corresponding to two of the species decreases until they disappear completely. Since the resonances corresponding to the native dimer are lost at similar concentration to the other species, it is likely that this third species also corresponds to the same previously unobserved dimer, especially given that the T419 residue forms a hydrogen bond with E409'. Moreover, the ESI-MS data showed the presence of no higher order oligomers. This alternative dimer is significantly populated at 500 μM. At this concentration there is a ratio of approximately 1:1:1 between the dimer, monomer and the alternative dimer is observed. This ratio was deduced from integrating the intensities of the three Trp side chain peaks, which were similar shapes. At lower concentrations (< 50 μM) peaks corresponding to the monomer are visible only. During the course of the dilution the intensity of the monomer peaks increases until it becomes the most dominant species.

Data recorded for the ^{15}N-T414A and ^{15}N-T414K p62 UBA mutants showed spectra that were highly similar to those previously recorded for wild type p62 UBA at similar concentrations. Moreover, the two mutant spectra correlated well to each other highlighting that the different mutants possess a similar tertiary structure to one another. At high concentration, a single species corresponding to the dimeric form, is observed. The peaks corresponding to the dimeric species for both mutants are well dispersed, indicative of a folded protein. At low concentration the spectrum becomes less clear. The dispersion of peaks becomes narrower, with most peaks between 7 and 8 ppm in the ^{1}H direct dimension. A narrow distribution of peaks would suggest partial unfolding of the protein. Partial unfolding of these mutants was initially indicated by larger solvent exposed surface areas relative to wild type p62 UBA in ESI-MS dilution studies. There is also evidence for protein degradation as highlighted by the poor definition of the peaks in the spectrum. Based on this information, it would appear that the monomer has been significantly destabilized, potentially as a result of partial unfolding, consistent with the ESI-MS observations for these mutants.



A)^{15}N-T414A B)^{15}N-T414K



C)^{15}N-T419K D)^{15}N-wild type p62 UBA

Figure 6.12 – High concentration (500 μM) HSQC spectra for the p62 UBA weak dimer mutants. (A) T414A p62 UBA, B) T414K p62 UBA, C) T419K p62 UBA and D) wild type p62 UBA. All spectra were recorded in25 mM potassium phosphate, 25 mM NaCl, pH 7. For wild type p62 UBA at 500 μM resonances corresponding to the dimer only are observed in the spectrum, as indicated by 1 Trp side chain peak in the bottom left corner of the spectrum. Similar to wild type p62 UBA, the spectra for theT414A and T414K mutants shows resonances for the dimer only. The T419K spectrum is more complex and reveals 3 species in equilibrium (as shown by the 3 Trp side chain peaks). This highlights how the equilibrium has shifted towards the monomer as resonances for both the monomer and dimer are visible at 500 μM. The third species corresponds to an alternative dimer.

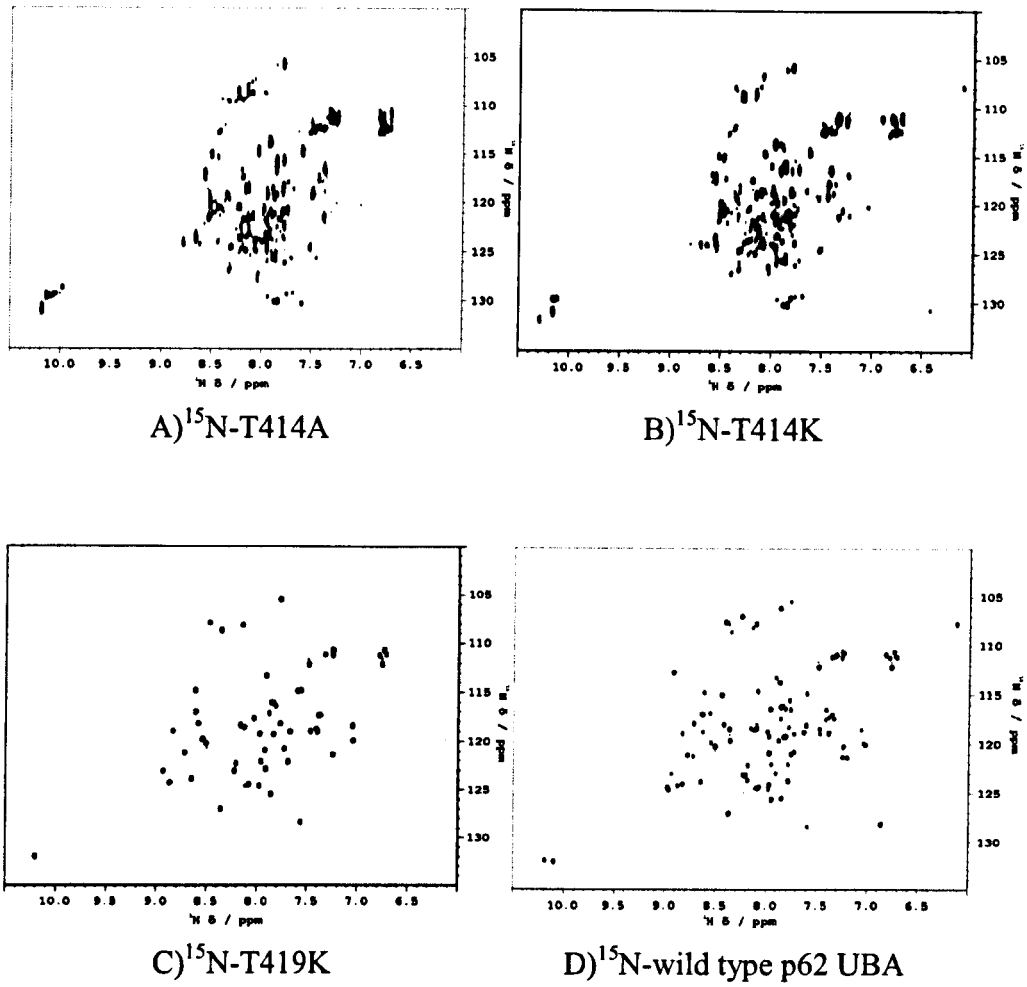


Figure 6.13 – Low concentration (10 μM) HSQC spectra for the weak dimer mutants A) T414A p62 UBA, B) T414K p62 UBA, C) T419K p62 UBA and D) wild type p62. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7. At 10 μM wild type p62 UBA, T414A p62 UBA and T414K p62 UBA show resonances for 2 species, the monomer and dimer. The 2 species are shown by the 2 Trp side chain peaks in the bottom left corner of the spectrum. A reduced stability is observed for the T414A and T414K mutants the quality of the spectra is low. The T419K spectrum shows resonances for a single species, the monomer. At this concentration the 2 dimer species have dissociated to the monomeric form. The T419K spectrum highlights how far the equilibrium has been shifted in the T419K mutant.

6.2.2.5 Low concentration NMR on the T419K monomer

The T419K mutant was the only mutant which showed the presence of the monomer at high concentrations indicating a substantial shift in the equilibrium towards the monomer. Since a fully monomeric NMR spectrum was recorded at concentrations below 50 μM investigations into the flexibility of the monomer could be conducted. There is some speculation about the flexibility of the monomer because in this form an extra half turn is formed by helix 3. No structural work on the monomer in isolation has been completed because the monomer is so lowly populated at concentrations required for NMR. The T419K mutant has shifted the equilibrium significantly enough to overcome this problem. In order to study the flexibility of the T419K monomer ^{15}N-heteronuclear NOE experiments were conducted. In order to calculate ^{1}H-^{15}N NOE values from the data, the interleaved ^{1}H-^{15}N HSQC experiments need to be separated and fully assigned.

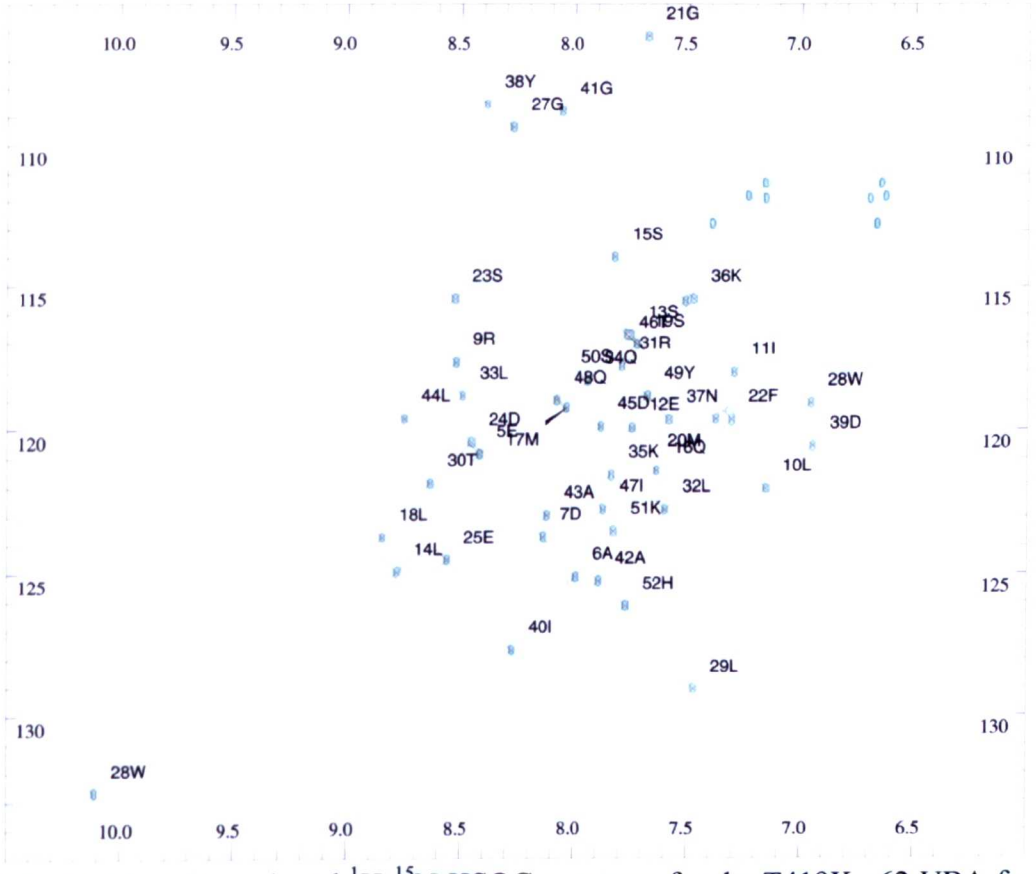


Figure 6.14 – The assigned ^{1}H-^{15}N HSQC spectrum for the T419K p62 UBA free monomer. The assignment numbers are based on the residue number within the UBA construct plus the N-terminal GS extension which remains after thrombin cleavage. The data was recorded on a 25 μM sample in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K.

The assignment of the T419K monomer was completed to approximately 90 % using previous assignments for the wild type monomer despite numerous residues in helix 2 showing significant changes in chemical shift. Most of the shifts were small enough to transfer the assignments across from the wild type protein. Resonances corresponding to Leu416, Leu417, Gln418 and Lys419 (the mutated residue) showed substantial chemical shift deviations and required further NMR experiments in order to complete the assignment. These residues were assigned using 2D NOESY and TOCSY experiments. Using an 800 MHz spectrometer and a TXI cryoprobe, NOESY and TOCSY experiments were conducted on a 25 μM sample. This concentration ensured that only resonances belonging to the monomer were detected. The fully assigned spectrum is shown in figure 6.14.

Once the T419K monomer had been assigned a ^{1}H-^{15}N heteronuclear NOE experiment was recorded on the T419K monomer using a 25 μM sample. These experiments are commonly used to probe the flexibilities of proteins. The experiment was recorded as two interleaved ^{1}H-^{15}N HSQCs, one in the presence of proton saturation and one without. The two spectra were separated and ^{1}H-^{15}N heteronuclear NOE values calculated by taking the ratio of ^{15}N signal intensities recorded in each spectrum (figure 6.15). In both cases weak signal intensities were recorded for all peaks. This is caused from the combined effects of low sample concentration as well as the weakened sensitivity of the heteronuclear NOE experiment. The residues with the largest ^{15}N-heteronuclear NOE values are found in the most rigid parts of the molecule. Negative values can be produced from completely unstructured parts of the protein as these regions are conformationally flexible. Heteronuclear NOE values ranging between 0 and 0.6 generally indicate relatively dynamic regions of the protein.

The T419K monomer displays heteronuclear NOE values which are very close to 0.6 for most residues. An average value of 0.62 was calculated. These values indicate the presence of an ordered structure which still exhibits some flexibility. The mutated residue Lys419 is the most flexible residue with a heteronuclear NOE value of 1.03. Gly411 also has a high NOE value of 0.87. The heteronuclear NOE values of both these residues are greater than the theoretical maximum value of 0.82[244], but in practice errors can occur when measuring signal intensities, creating larger NOE values. Gly residues are usually quite dynamic. Gly411 is

found in the loop 1 region and it is not surprising that this residue exhibits conformational flexibility.

Glu389, Lys435 and His436 were all only detected in the spectrum with proton saturation. These peaks have broadened significantly and are unable to be detected in the spectrum without proton saturation. This indicates these residues are highly flexible, consistent with these residues being located close to the termini of the T419K. The heteronuclear NOE values produced for these three residues were inaccurate and were therefore excluded from the analysis. In addition to this the heteronuclear NOE values for residues Ile431 to Ser434 are all positive, in agreement with the wild type p62 UBA monomer, in which helix 3 extends by an extra half turn. If residues Ile431 onwards exhibited strongly negative signals, it would be expected that these were forming part of the flexible C-terminus.

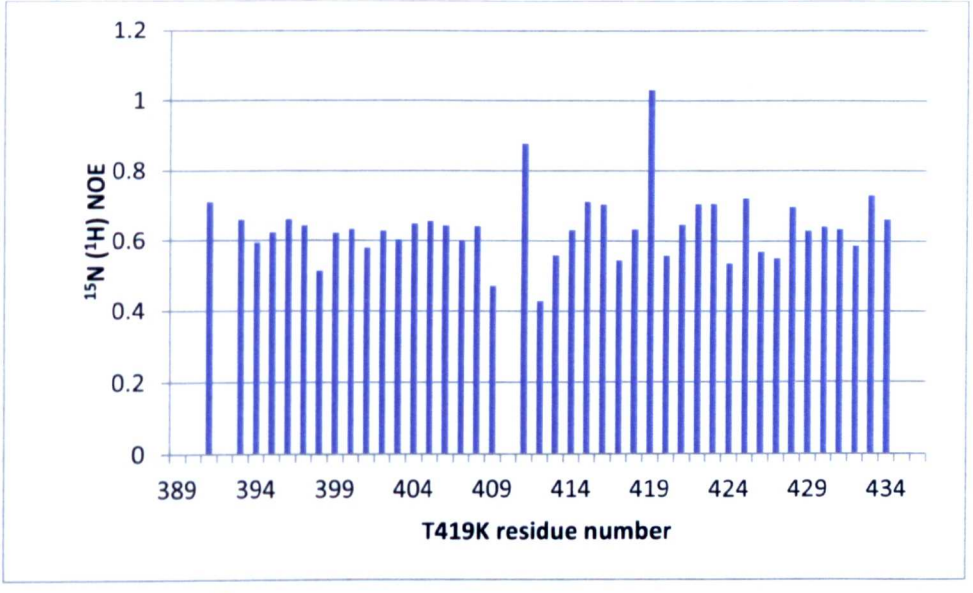


Figure 6.15 – The ^{15}N-heteronuclear NOE values for the T419K p62 UBA free monomer. The ^{15}N-heteronuclear NOE experiments were recorded on a 25 μM sample in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K. The values were calculated using CCPNMR version 2.1.2 software. Values for Glu389, Lys435 and His436 were inaccurate and were subsequently excluded from the analysis. Most of the values are close to 0.6 indicating that the monomer shows limited structural flexibility.

6.2.2 NMR binding studies on the p62 UBA weak dimer mutants

6.2.2.1 ^{15}N-T419K p62 UBA titrations

The dilution NMR study indicated a significant population of the monomer at higher concentrations. In order to investigate if the shift in equilibrium towards the monomer has an effect on binding affinity an NMR titration whereby unlabelled ubiquitin was titrated into ^{15}N-T419K p62 UBA was performed. At high concentration peaks corresponding to the monomer could be visualised. The monomer peaks could then be tracked from the free to the bound form to calculate a K_d for binding.

Peaks corresponding to the free monomer were assigned in the free ^{15}N-T419K UBA 500 μm spectrum. The assigned peaks exhibited mixture of strengths in signal intensity. Some were quite weak; whereas, others were quite strong. It was assumed that the weak peaks corresponded to the monomer only. The strong intensity of the other peaks was likely to be caused by the presence of either or both of the dimers which produce resonances which overlap with the monomer. Since the chemical shift rather than peak intensity was used for analysis, this was not considered to be problematic.

Up to 2 mM unlabelled ubiquitin was titrated into a 500 μM ^{15}N-T419K p62 UBA sample. Concentration dependant chemical shifts were observed upon the addition of ubiquitin consistent with the binding of the UBA to ubiquitin. A mixture of slow and fast chemical exchange regimes were seen, similar to the titration between wild type p62 UBA and ubiquitin, suggesting that a similar mechanism of dimer dissociation is occurring prior to ubiquitin binding by the monomer, despite the presence of a second uncharacterised dimer. During the titration peaks corresponding to the two dimeric species decreased in intensity whilst the peaks corresponding to the free monomer increased in intensity. Whereas, fast exchange was observed between the free and bound forms of the monomer. Saturation was observed at 500 μM, a 1:1 molar ratio. Wild type p62 UBA doesn't reach saturation until 2 mM ubiquitin, a 1:4 molar ratio; therefore, the T419K mutant shows a significant increase in the binding affinity.

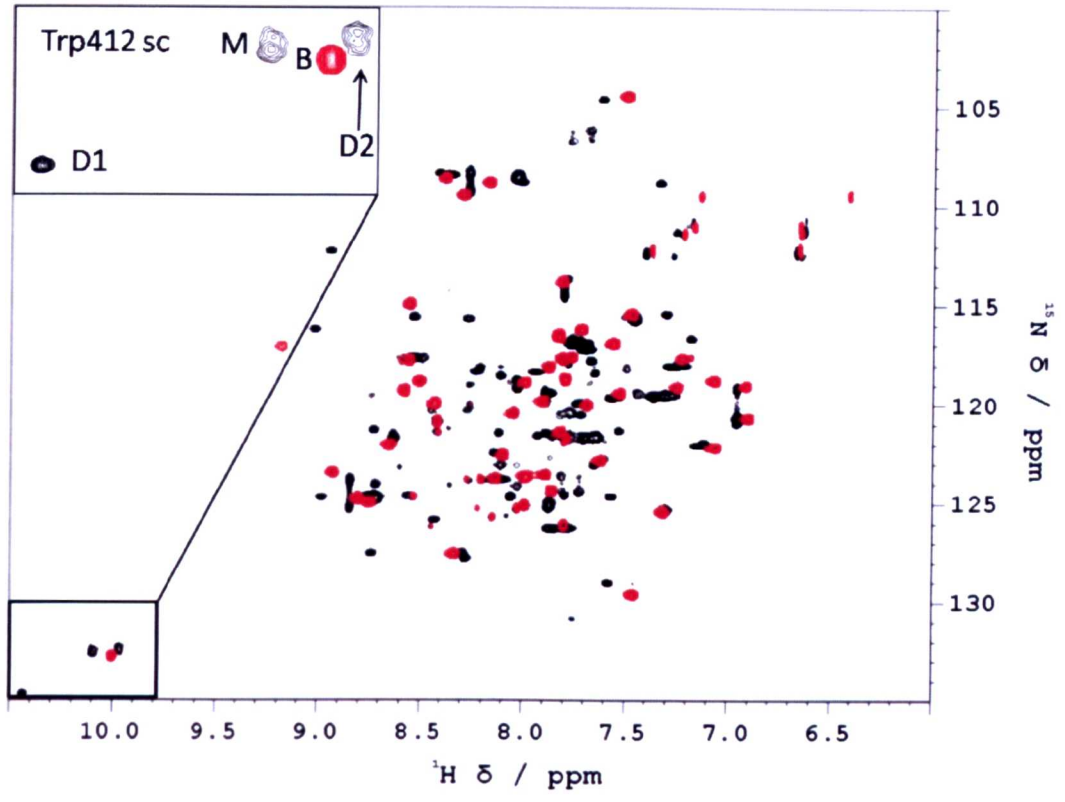


Figure 6.16 – ^{1}H-^{15}N HSCQ spectrum for the free ^{15}N-T419K p62 UBA at 500 μM (black) overlaid with the bound to ubiquitin spectrum at 1:4 molar ratio (red). The 3 species in the free spectrum move to a single bound species upon the addition of ubiquitin. The Trp side chain region is enlarged to visualise the 3 species (dimer1 'D1', dimer 2 'D2' and monomer 'M') in the free spectrum and the 1 species in bound spectrum (bound 'B'). The spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH7 and at 298 K.

The three species observed in the free ^{1}H-^{15}N HSQC spectrum all move to a single species, the bound monomer form (figure 6.16). The bound T419K monomer form has resonances with chemical shifts which are nearly identical to those previously observed for wild type. Small CSPs were observed for residues Trp412, Leu413, Arg415, Leu416, Leu417, Gln418, Asn421 and Tyr422. All of these residues were able to be assigned as they had not moved significantly. Only Lys419 and Lys420 were unable to be identified from the bound spectrum without the aid of the dilution data which displayed the free monomer form. However, these residues were able to be tracked from the free to the bound monomer form, thus completing the assignment of the bound form. Unfortunately the presence of the third species produced complicated spectra which precluded the ability to assign the two dimer species without ^{13}C and ^{15}N labelling.

Changes in the chemical shift between the free and bound forms were plotted against residue number to identify which residues showed the greatest perturbations during the titration (figure 6.17). The pattern of changes in chemical shift correlated well to the data obtained for the wild type monomer, highlighting the MGF motif in loop 1 and residues in helix 3 as the biggest movers. The similarity between the bound form of the T419K and wild type p62 UBA confirms that the mechanism of binding had not been altered by the mutation and that the same bound form was reached. The alternative dimer would appear to maintain the ability to bind to ubiquitin. The faster saturation was the direct result of an enhanced affinity interaction. Since the T419K has a significantly higher population of monomer at 500 μM compared to wild type protein it was expected that an increased affinity for ubiquitin would occur.

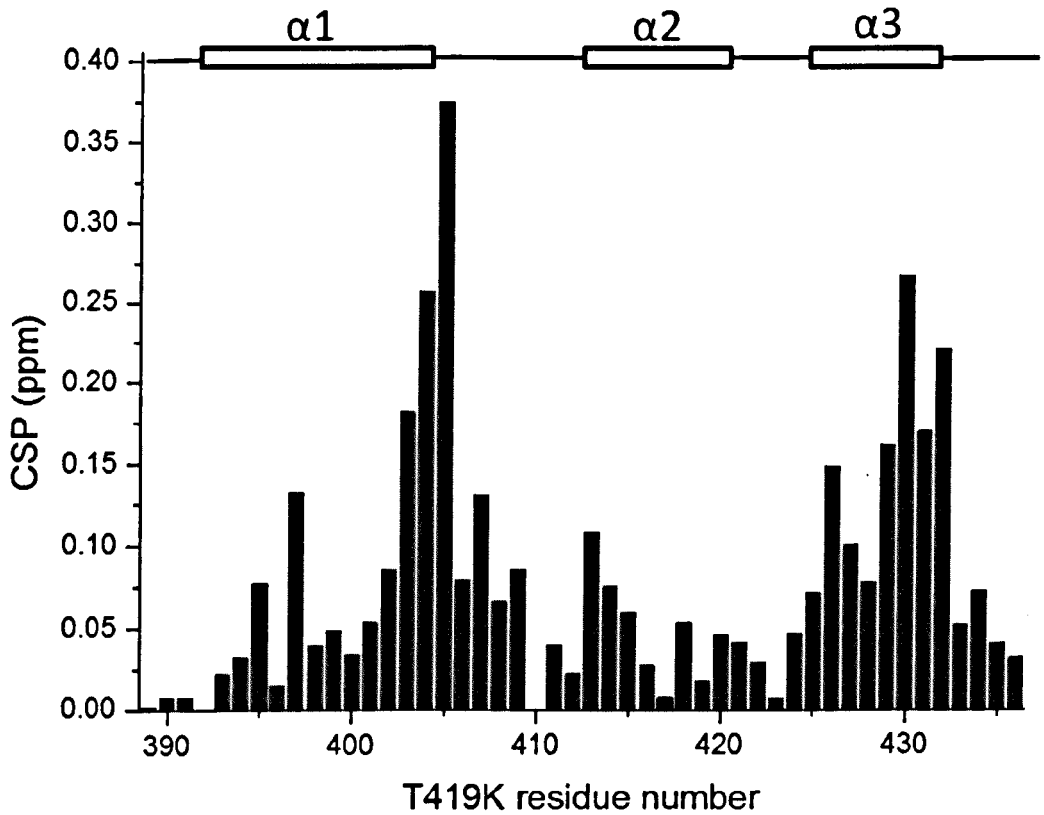


Figure 6.17 – Residue specific CSP for the backbone NH groups for the ^{15}N-T419K p62 UBA monomer between the free and bound forms. The gaps correspond to residues which were unable to be assigned. The largest CSPS are seen for Met404 and Gly405 in the conserved MGF motif in loop 1 and residues in helix 3. These large CSPs highlight the T419K mutant binds to ubiquitin using a similar mechanism to wild type p62 UBA.

By globally fitting the residues with large changes in chemical shift between the free and bound forms (CSP >0.05 ppm) to a simple 1:1 binding model using IGOR Pro an apparent K_d could be produced. A total of 24 curves were subsequently fitted to produce a K_d value of 185.9 ± 8.7 μM. The curves fitted well to a 1:1 binding model. The K_d value is best described as the observed K_d (K_{obs}) because the fit does not take into consideration the competitive processes of dimerisation and binding. By using the relationship between K_{obs} and K_{dim} (equation 12, chapter 2), which takes into account the two competing equilibria, a more accurate K_d value can be calculated. A new K_d value of 175.1 ± 20.4 μM was therefore calculated using the K_{dim} value 165 ± 30 μM determined from the ITC. However, as mentioned earlier the fit produced for the T419K mutant was poor. The wild type p62 UBA monomer was previously shown to have a K_d of 40 ± 10 μM. It is clear from the titration data that the T419K has a tighter interaction with ubiquitin because saturation occurs 4-fold quicker than it does for wild type. The K_d for binding calculated using K_{obs} and K_{dim} is likely to have been seriously affected by the inaccurate K_{dim} value.

The residues which showed a CSP of more than 0.1 ppm were considered to be significant movers. These residues were plotted onto the surface of the wild type bound monomeric UBA and used to identify the binding surface (figure 6.18, PDB ID 2JY8). The darkest red colour indicated residues with the greatest the change in chemical shift. The MGF loop which protrudes from the surface of the structure is coloured red, highlighting this as the interacting surface. It is therefore likely that the T419K monomer maintains the interaction between the MGF loop and the Ile44/Val70 hydrophobic patch.

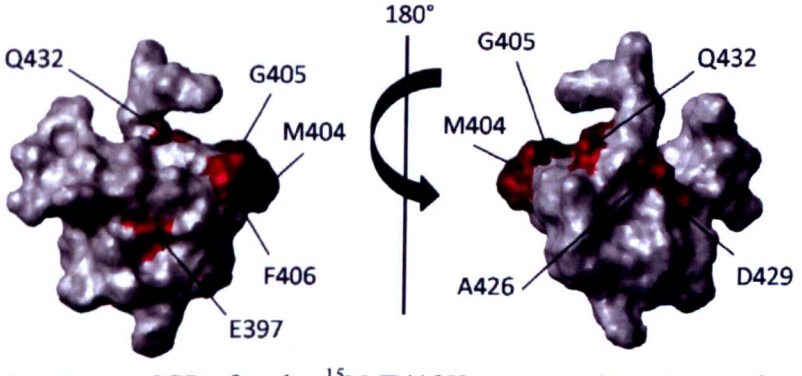


Figure 6.18 – Large CSPs for the ^{15}N-T419K mutant plotted onto the surface of the bound monomeric UBA (PDB ID 2JY8). The darker the red, the greater the perturbation.

6.2.2.2 ^{15}N-T414A and ^{15}N-T414K p62 UBA titrations

In order to investigate if the T414A and T414K p62 UBA mutants also showed an enhanced affinity for ubiquitin, binding studies using NMR spectroscopy were completed. These mutants did not have the added complication of an alternative dimer; however, even at concentrations as high as 500 μM some additional weak peaks were observed in the free ^{1}H-^{15}N HSQC for both the ^{15}N-T414A and ^{15}N-T414K mutants when the contour levels were increased. These peaks were believed to correspond to the free monomer as the biophysical data indicated that the equilibrium had shifted towards the monomer. The extra resonances were not assigned as an incomplete set of monomer resonances were observed at high concentrations. Unfortunately these mutants were too unstable to conduct low concentration NMR titrations whereby a full set of peaks corresponding to the free monomer can be visualised.

The ^{1}H-^{15}N HSQC spectra for the free ^{15}N-T414A and ^{15}N-T414K mutants showed peaks with a similar pattern of dispersion to wild type p62 UBA. The spectra for the ^{15}N-T414A and ^{15}N-T414K UBAs were near identical, although a considerable number of resonances had chemical shifts which had deviated from the wild type (figure 6.19). For some resonances the assignments could be transferred from the wild type to the mutant, but others showed changes which prevented unambiguous assignment. The unassigned residues showed larger changes in chemical shift compared to wild type p62 UBA and were mainly clustered in the vicinity of the mutation site in loop1 and helix 2. The unassigned residues were assigned using NOESY and TOCSY data recorded on a 1 mM sample of the T414A mutant. Since the chemical shifts were so similar for the two mutants, NOESY and TOCSY spectra were only recorded for the T414A mutant.

The ^{1}H-^{15}N HSQC spectra for the bound forms unsurprisingly also showed some resonances with considerable changes in chemical shift compared to wild type p62 UBA. The resonances which were different this time were located in helix 1 as well as loop1 and helix 2. Some of these residues were able to be assigned using the assignments previously reported for wild type p62 as they were located in isolated regions of the spectrum. For residues which were unable to be assigned a ^{15}N-HSQC-NOESY and ^{15}N-HSQC-TOCSY were run on a fully bound 125 μM ^{15}N-T414A sample (1:2 molar ratio ^{15}N-T414A:unlabelled ubiquitin). Again the

bound form of both the T414A and T414K mutants were so similar that the assignments could be transferred between the spectra.

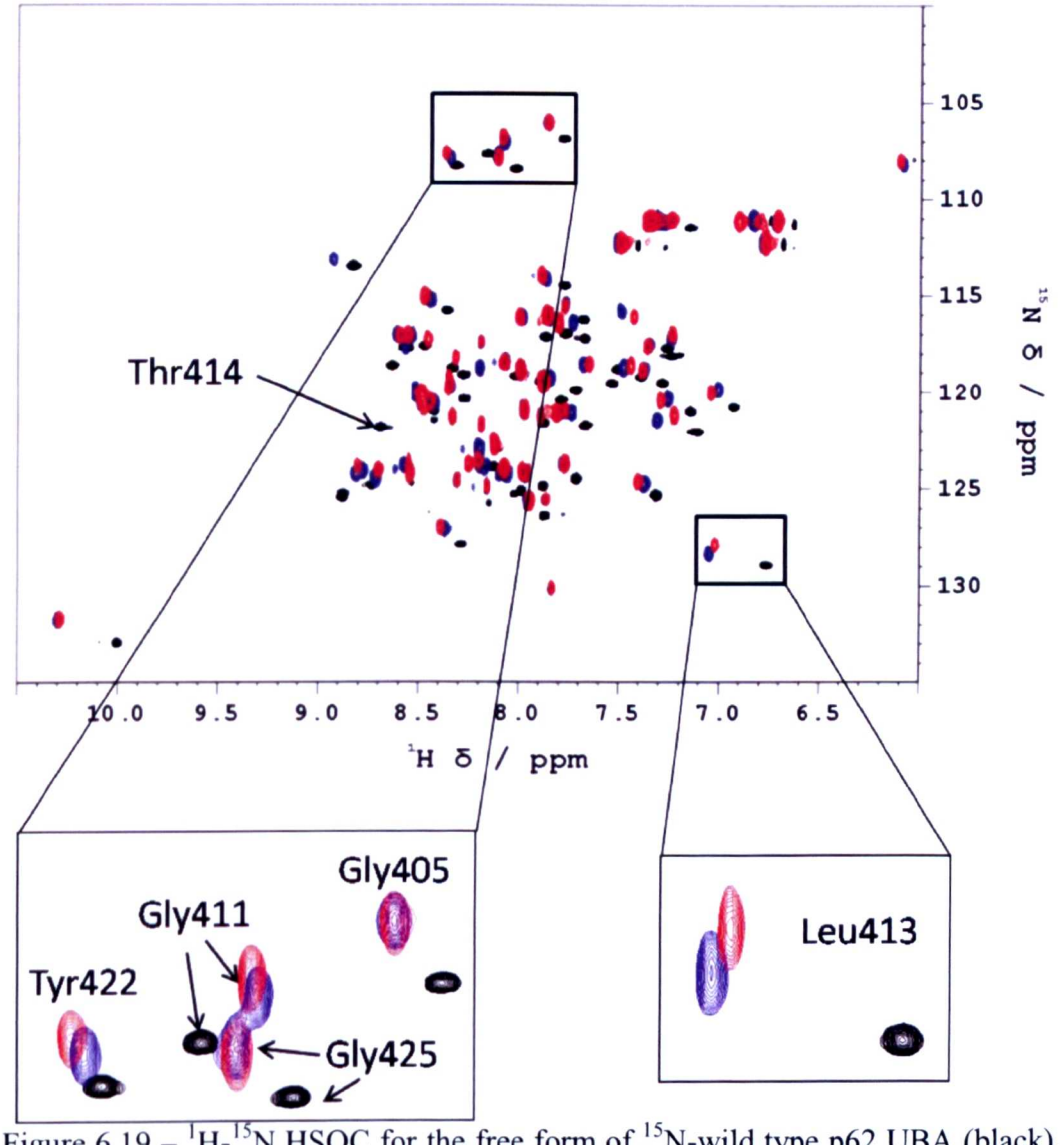


Figure 6.19 – ^{1}H-^{15}N HSQC for the free form of ^{15}N-wild type p62 UBA (black), ^{15}N-T414A p62 UBA (blue) and ^{15}N-T414K p62 UBA (red) at 250 μM. Regions of the spectrum where resonances for the two mutants show changes in chemical shift compared to wild type are expanded. The mutated residue, Thr414, is also labelled. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH7 and at 298 K.

Titration studies whereby unlabelled ubiquitin was titrated into ^{15}N-T414A and ^{15}N-T414K were conducted. Up to 1 mM unlabelled ubiquitin was added to 250 μM ^{15}N-T414A and ^{15}N-T414K, as separate experiments (figures 6.20 and 6.21). At 250 μM, the weak additional peaks had increased in intensity, again consistent with them corresponding to the monomer. Concentration dependent changes in chemical shift were observed in both titrations, consistent with a binding event. Saturation occurred after the addition of 500 μM ubiquitin (1:2 molar ratio mutant UBA:ubiquitin) in both cases. The mutants therefore showed an increased affinity for ubiquitin, as the wild type saturated at a 1:4 molar ratio (UBA:ubiquitin). However, these mutants do not show as strong interaction as the T419K mutant. A mixture of fast and slow exchange was observed consistent with the data obtained for the wild type p62 UBA. The two different chemical exchange regimes correspond to two separate events, the dimer dissociation and ubiquitin binding. The peaks corresponding to the dimer decrease in intensity whilst the peaks corresponding to the monomer increase in intensity. Concurrently the peaks corresponding to the monomer move in fast exchange to the bound form.

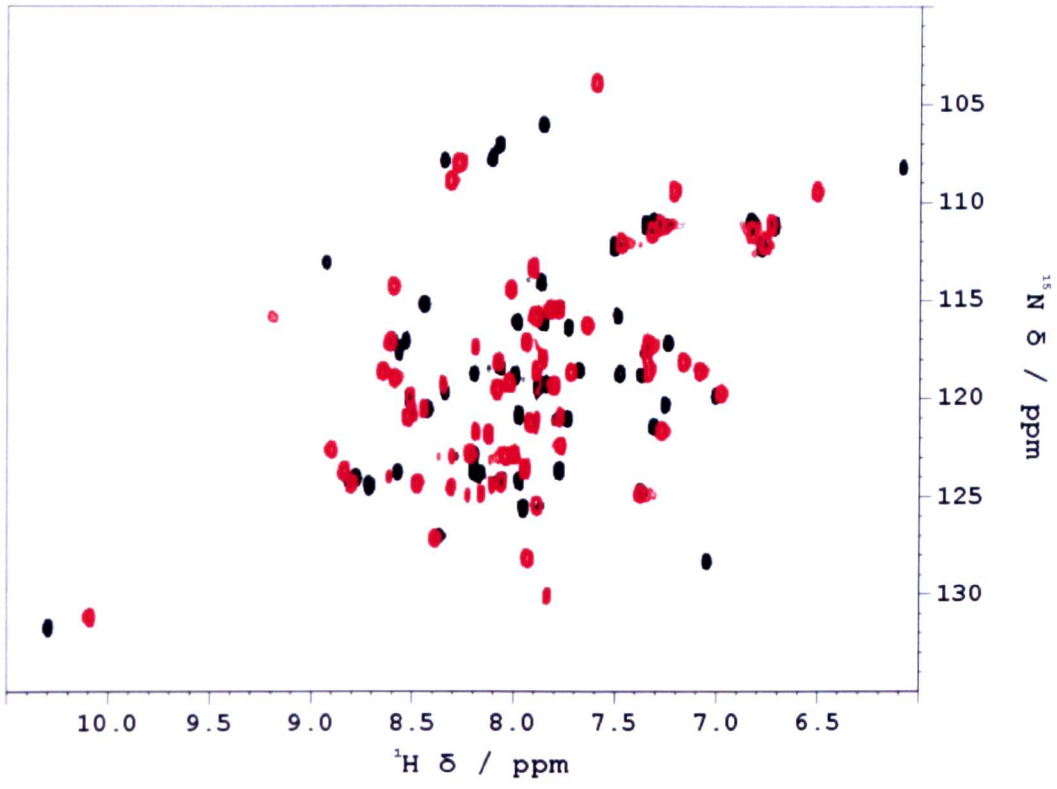


Figure 6.20 - ^{1}H-^{15}N HSQC spectrum of the free ^{15}N-T414A at 250 μM (black) overlaid with the bound to ubiquitin spectrum at 1:4 molar ratio (red). The spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH7 and at 298 K.

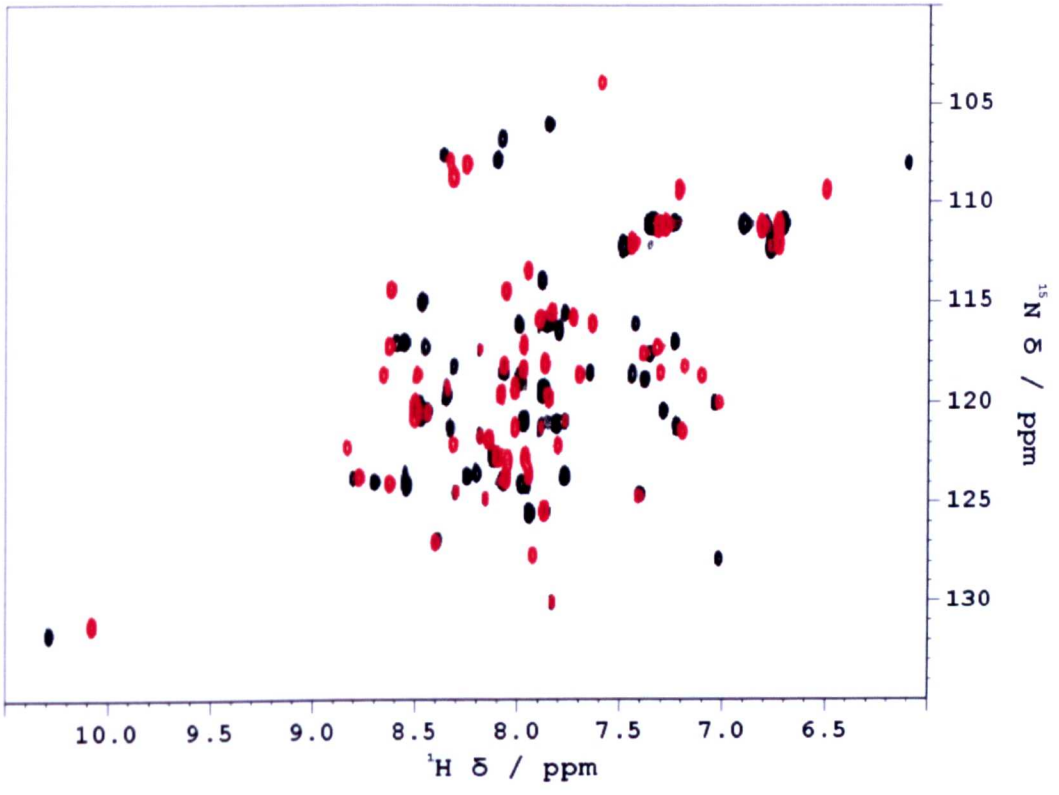


Figure 6.21 - ^{1}H-^{15}N HSQC spectrum of the free ^{15}N-T414K at 250 μM (black) overlaid with the bound to ubiquitin spectrum at 1:4 molar ratio (red). The spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH7 and at 298 K.

By plotting the changes in chemical shift between the free and bound forms on a per residue basis, the regions of the molecule that have been affected were identified (figure 6.22). The correlation between the CSPs observed by the T414A and T414K datasets was near identical. At this concentration the combined effects of the transition from dimer to monomer as well as binding by the free monomer are visualised, resulting in large CSPs. An accurate determination of the K_d for binding was therefore not obtainable from this data. However, the CSP pattern is in good agreement with the CSPs previously obtained for the wild type p62 UBA at high concentrations, suggesting the mechanism of binding is the same for the wild type and the two mutants. Residues with large CSPs in helix 2 and the C-terminal of helix 3 are caused from the dimer dissociation; whereas, large CSPs in loop1 and helix 3 are caused from ubiquitin binding.

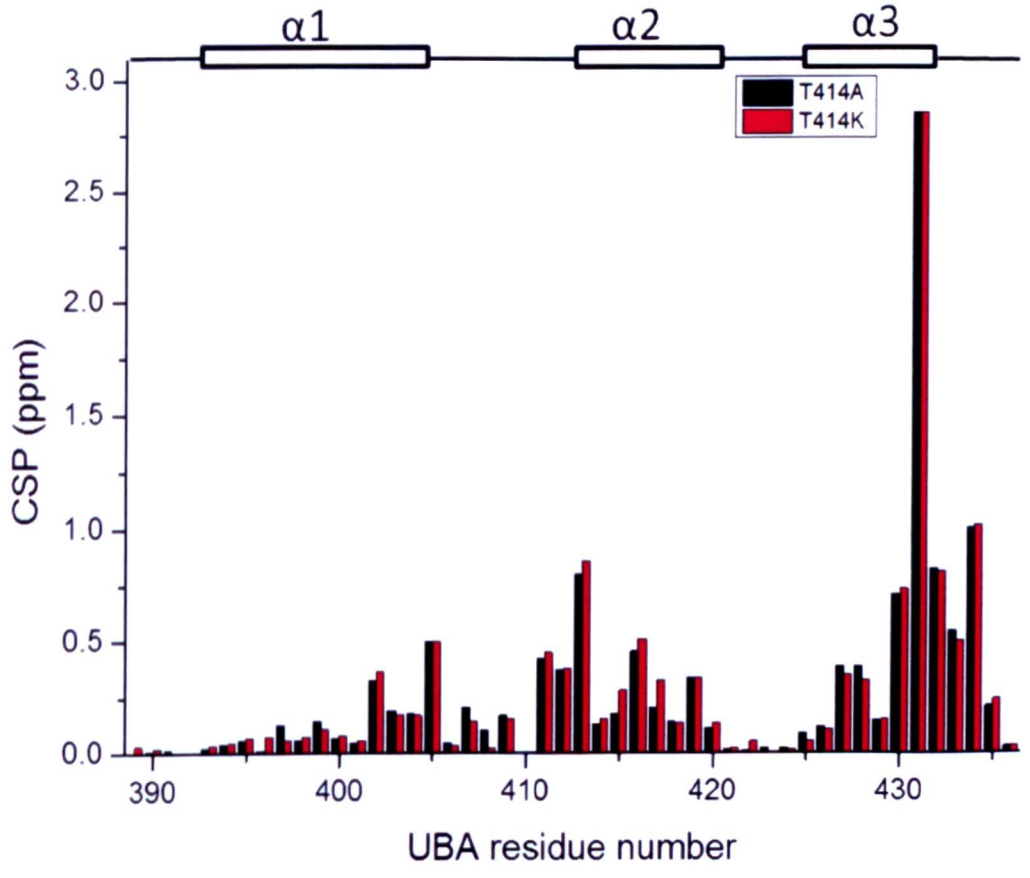


Figure 6.22 – Residue specific CSPs for the T414A and T414K p62 UBA mutants upon ubiquitin binding. Large CSPs are exhibited by residues in helix 2 and helix 3. Helix 2 corresponds to dimer dissociation whereas the effects observed in helix 3 are exerted by both dimer dissociation and ubiquitin binding.

Residues which showed a CSP of more than 0.05 ppm were considered to be significant movers. These residues were plotted onto the surface of the wild type monomeric UBA in the bound form (figure 6.23, PDB ID 2JY8). A large proportion of the molecule is affected due to the two processes. Residues which have been coloured red respresent regions of the molecule which have been affected by binding, with the darker red colour corresponding to the bigger CSPs. Residues which have turned black in colour indicated the residues with the biggest CSPs. Regions of the molecule which have turned black in colour correspond to residues in helix 3 which are affected by both dimer dissociation and ubiquitin binding. Residues in helix 2 are coloured the darkest reds, with Leu413 also coloured black. These residues have been perturbed from dimer dissociation. The MGF loop is also coloured red indicating binding to ubiquitin has occurred using these residues.

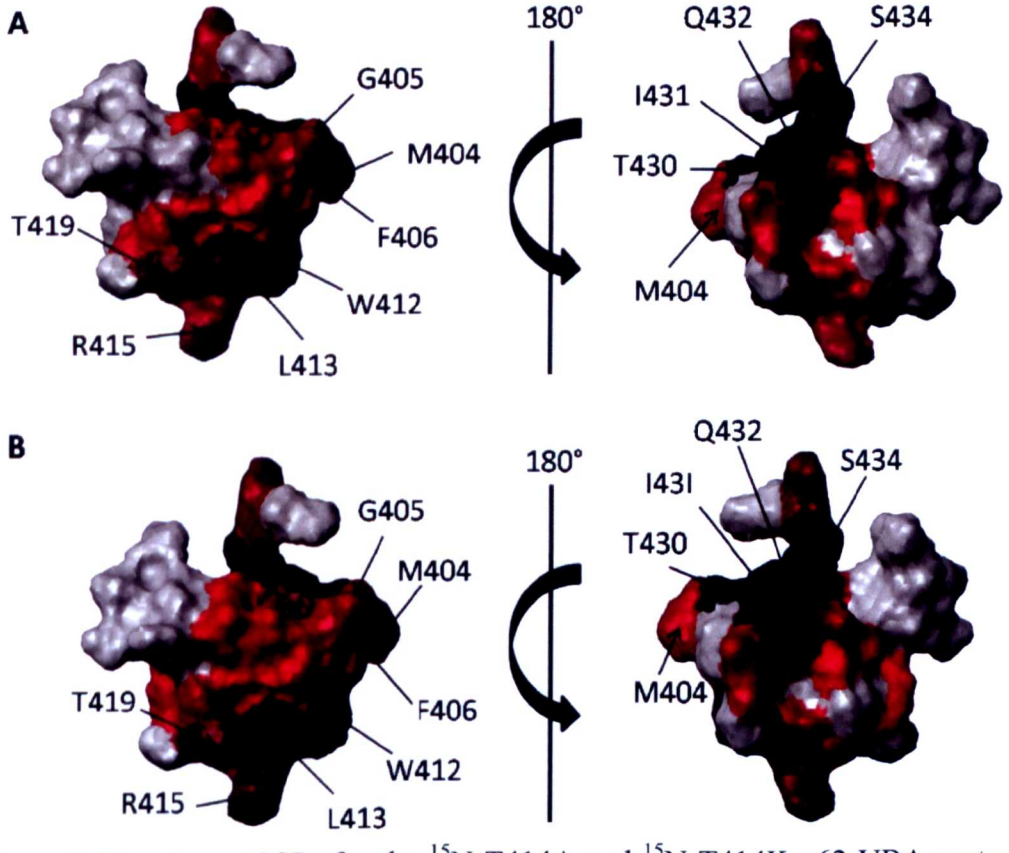


Figure 6.23 – Large CSPs for the ^{15}N-T414A and ^{15}N-T414K p62 UBA mutants plotted onto the surface of the bound monomeric UBA (PDB ID 2JY8). The darker the red colour, the greater the perturbation. The observed CSPs are caused by both dimer dissociation and ubiquitin binding. Residues T430, I431, Q432 and S434 are coloured black and correspond to residues which have the greatest CSPs. These residues are affected by both dimer dissociation and ubiquitin binding.

6.3 Discussion

The competitive equilibrium between dimerisation and ubiquitin binding represents a highly sophisticated control mechanism, which is able to regulate the affinity the UBA has for ubiquitin. The p62 UBA dimerisation interface is primarily composed of residues in helix 2, although residues at the C-terminal end of helix 3 also play a part in maintaining dimer stability. Upon dissociation to the monomer an extension in helix 3 has been proposed which is able to prevent dimerisation, as well as facilitate the interaction with ubiquitin. The dimer therefore has an inhibitory role which opposes the effects of ubiquitin binding. Understanding the relationship between these two factors, and potentially others, is crucial to understanding the molecular basis of Paget's disease.

Despite the formation of significantly weakened dimers, none of the mutants studied in this chapter were able to form solely monomeric structures. The only mutant which permitted the study of the monomer was the T419K mutant, which was solely monomeric at concentrations below 50 μM. The study by Isogai *et al* which produced the crystal structure for the p62 UBA dimer also generated mutants to probe the monomer-dimer equilibrium. They used an approach which targeted the key hydrophobic interactions at the interface and were able to produce monomeric UBAs at 100 μM for six of the seven mutants[131]. Although targeting the hydrophobic residues at the binding interface showed a greater shift in the monomer-dimer equilibrium, these mutants also formed dimers at higher concentrations. Taken together, it is clear that the dimer is the preferred conformation at physiological concentrations.

In addition to this, the UBA dimer might also have a role in determining the life span of the full length p62 protein. The p62 is not normally degraded by the proteasome; however, deletion of the C-terminal eight residues prevented dimer formation but interestingly also promoted proteosomal degradation[131]. A protective role for C terminal UBA domains has been demonstrated by UBA in the Rad23 and Dsk2 proteins[144], although these domains did not form dimer structures. These UBA monomers were able to prevent generation of initiation sites for degradation, which are required for interaction with the unfolding machinery of the proteasome. This is a feature which could be shared by the p62

UBA dimer, or alternatively the dimer might be prevented from binding to the proteasome due to steric clashes linked to dimer formation.

When the p62 UBA dimer was discovered only a few examples of UBA dimers had been reported. The UBA domains of the c-Cbl and Cbl-b proteins and the doublesex UBA from *Drosophila melanogaster* had all been previously shown to exist as dimers[234,235,237,245]. Also since the discovery of the p62 UBA dimer, the structure of the XIAP UBA dimer has also been reported[236]. Interestingly all of these proteins form structurally distinct UBA dimers, with the C2 symmetry only exhibited by the p62 UBA dimer. And rather more surprisingly not all of these UBA dimers appear to regulate ubiquitin binding. No other UBA dimer possesses overlapping dimerisation and ubiquitin binding interfaces. This would suggest that the p62 UBA dimer has specifically evolved under pressure to regulate function.

Many UBA-ubiquitin interactions have been shown to be weak[101,54], although examples of higher affinity interactions have also been reported[98,99]. Of the UBA dimers a range of affinities has also been determined. The Cbl-b UBA has been reported to have a higher affinity interaction (K_d 53 ±3 μM)[245]; whereas, the XIAP UBA shows a weaker interaction (K_d 249 ± 19 μM)[236]. Many UBAs have been shown to bind to different types of polyubiquitin chains, with certain chains preferred over others. It would therefore appear the UBAs are able to intrinsically regulate their binding affinities dependent on the polyubiquitin chains they are binding. The p62 UBA thus far exhibits no linkage preference, but does show enhanced binding to polyubiquitin chains over monoubiquitin in GST pull downs[170,171]. Due to the highly oligomeric nature of the p62 *in vivo*, it is likely that multiple UBAs are presented to polyubiquitin chains. This would suggest binding of the p62 is mediated by avidity effects, which favour the open conformation of Lys63 linked chains[93]. Avidity effects are likely to filter out the non-specific interactions and amplify the correct UBA-ubiquitin interactions. The UBA dimer however adds complexity to this model since high concentrations of UBA are likely to promote dimer formation and repress ubiquitin binding. The overall binding affinity of the UBA in oligomeric p62 also involves competitive effects between UBA dimerisation and avidity. It is therefore speculated that changes in affinity as a result of the formation of weaker dimers are amplified by avidity effects.

6.4 Conclusions

The main aim of this chapter was to probe the monomer-dimer equilibrium by disrupting or preventing dimer formation. By introducing mutations to helix 2, which forms the dimerisation interface, the structural integrity of the dimer and changes in the equilibrium were monitored using structural and biophysical techniques. The presence of the UBA dimer has an inhibitory effect on NF-κB activation by decreasing the affinity for ubiquitin[141]. By altering the strength of the dimer, the equilibrium can be shifted causing an increase in the population of the monomer, thereby increasing the affinity for ubiquitin. The mutants studied in this chapter produced a mixture of wild type like dimers and weak dimers. Of the eight designed mutants, seven were able to be successfully expressed and purified in C41 (DE3) *E. coli*. Four of the mutants formed dimers similar to wild type p62 UBA and three formed significantly weaker dimers. Only the Q418K p62 UBA mutant was unable to be expressed, suggesting that this mutation was the most detrimental to dimer structure and stability.

The R415A, Q418A, T419A and R415K mutants all produced dimers which were not too dissimilar from the wild type. All of these mutants exhibited small differences in thermal stability and the solvent exposed surface area relative to wild type, as shown by Far UV-CD and ESI-MS dilution studies. Moreover, the structural integrity of the dimer was not significantly altered as shown by ITC. The Q418A mutant was shown to have the biggest increase in solvent exposed surface area as well as displaying the biggest thermal destabilisation. Whereas, the T419A mutant was shown to have the weakest K_{dim} value. Most of the Ala mutants did not seem to significantly affect the monomer-dimer equilibrium, suggesting that these residues did not form critical interactions at the dimer interface. The data also suggests that the hydrophobic residues which form the dimer interface are the driving force in maintaining dimer stability. The only Lys mutant which retained properties close to wild type p62 UBA was the R415K mutant, which was not unsurprising as the mutant involves a like-for-like change in amino acid.

However, the T414A, T414K and T419K mutants all produced dimers weaker by at least an order of magnitude compared to wild type protein, as shown by the K_{dim} values calculated from ITC dilution studies. K_{dim} values of 141 ± 1 μM, 146 ± 1.5 μM and 165 ± 30 μM were recorded for the T414A, T414K and T419K mutants respectively; whereas, the K_{dim} for the wild type p62 UBA is 4.1 ± 0.6 μM. The far UV-CD data for the T414A and T414K mutants reveal the thermal stability of both the monomeric and dimeric species are significantly weaker than those reported for the wild type protein. The T414A and T414K mutants show a 14 K and 21 K decrease in dimer stability respectively. The monomer is destabilized to a lesser extent than the dimer, however a large decrease in observed compared to the wild type. Decreases of 8 K and 15 K for the T414A and T414K monomers were observed respectively. In addition, the T_m values of the both the monomer and dimer for these two mutants were shown to be indistinguishable, highlighting the dimer as the species most affected by the mutation. The far UV-CD scans highlighted weaker intensity bands at 208 and 222 nm for the T414A mutant relative to wild type p62 UBA, suggesting a decrease in the average helical content.

The dilution NMR studies confirmed the presence of a partially unfolded monomer at low concentration (10 μM) for the T414A and T414K mutants. Low concentration NMR spectra for these mutants showed peaks with poor resolution and dispersion. It is likely that the partial unfolding of these mutants is linked to their decreased thermal stabilities. Whereas, the dilution NMR experiments for the T419K mutant revealed the presence of a third species. This third species had been previously observed by the E409K and G410K mutants[141]. This additional species was thought to be an alternative dimer which despite forming a different structure maintained the C2 symmetry along the dimerisation interface. Although the fit used to produce the K_{dim} for the T419K was poor, the other biophysical data indicated that this mutant formed a weak dimer. The presence of the monomer at high concentrations clearly shows that the equilibrium has moved towards the monomer.

The affinity of the weak dimer mutants was also investigated. Binding studies using NMR spectroscopy revealed that these mutants had an increased affinity for ubiquitin. The T414A and T414K mutant saturated at a 1:2 molar ratio. These

mutants showed a 2 fold increase in affinity compared to the wild type p62 UBA. The CSPs observed between the free and bound forms of the T414A and T414K mutants showed a very strong correlation with each other and were highly similar to wild type. The T419K mutant saturated at a 1:1 molar ratio. This mutant showed the biggest increase in affinity, a 4 fold increase, compared to wild type. This is consistent with the increased population of monomer at high concentration (500 μM). There is an obvious correlation between dimer strength and affinity for ubiquitin. The weaker the dimer, the further the equilibrium has shifted towards the monomer, the higher the affinity for ubiquitin. This is also consistent with only the monomeric form capable of binding.

The T414A and T414K mutants highlight residue Thr414 as a critical amino acid in the stability of the p62 UBA dimer. Also, as predicted, the T419K mutation showed the ability to severely affect the strength of the UBA dimer. The hydrogen bond formed by the Thr419 with Glu409 can no longer be formed, significantly destabilizing the dimer. The affect is so drastic that an alternative dimer is formed by the mutant. This alternative dimer was also observed by the E409K and G410 mutants[141].

Taken together, a relationship between dimer stability and affinity for ubiquitin is observed. The data indicates that the equilibrium in the weak dimer mutants has shifted towards the unstable biologically active monomer causing an increase in the population of monomer at equilibrium. The larger population of monomer in turn raises the affinity for ubiquitn. Since no solely monomeric species were observed it is concluded that the dimer is the preferred oligomerisation state for the p62 UBA dimer. The dimers regulatory role in NF-κB signalling is evident and ablation of dimer formation would be severely detrimental to this pathway. However, by mutating residues along the dimerisation interface we can produce weak dimers with a higher affinity for ubiquitin.

6.5 Future work

The results from this chapter as well as previous work on the wild type p62 UBA and E409K and G410K p62 UBA mutants clearly indicate a preference for the dimeric form of the UBA. Taken together, these results indicate that the monomeric form is weakly populated, as dissociation only occurs in the presence of ubiquitin and drives the formation of the complex. In order to gain a greater insight into the p62 UBA monomer by obliterating dimer formation completely a strategy involving multiple mutations along the dimerisation interface could be employed. Double, triple or even quadruple mutations at key interacting residues along the interface which target both the hydrophobic interactions and solvent exposed residues are more likely to prevent the dimer forming at all. Hydrophobic residues in helix 3 should also be considered when designing such mutants.

The T419K monomer could have been studied further. A K_d for binding could have been calculated by conducting binding experiments using ITC or ESI-MS. However, the mutations by Isogai *et al* were published before these experiments could be conducted. It was therefore deemed unnecessary to continue investigating the T419K mutant.

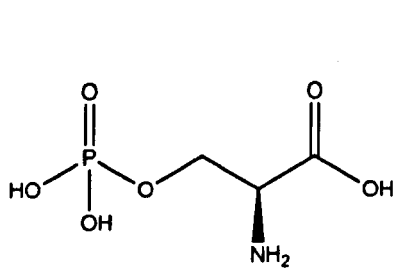
It would also be worthwhile inserting the T419K mutation or any of the mutations outlined by Isogai *et al* into some of the PDB mutants, such as P392L, M404V and G425R and monitoring the effects on ubiquitin binding. The P392L mutation is the most common mutation found in patients with Paget's disease, the critical MGF motif has been mutated in the M404V mutant and the G425R mutation is associated with a complete loss of ubiquitin. By studying the monomeric form of the UBA the role that dimerisation has in the various PDB mutants can clearly be established. Experiments using longer polyubiquitin chains and chains of various linkage could also be explored using the monomeric PDB mutants, as these are more likely to be the natural ligands. The different PDB mutants could show differences in affinity for certain linkages and could provide an insight into the onset of the disease.

7.0 Investigating phosphorylation of the p62 UBA

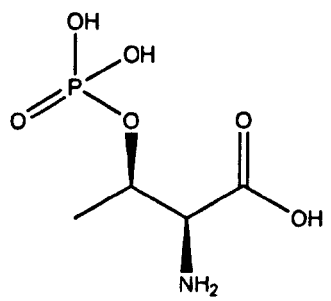
7.1 Introduction

7.1.1 Phosphorylation

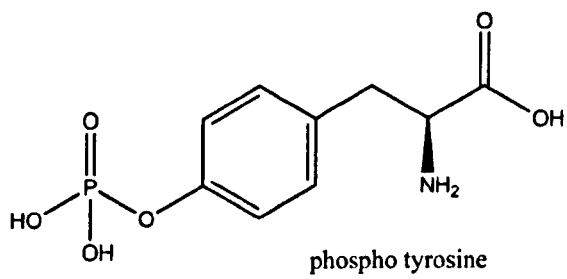
Protein phosphorylation involves the addition of phosphate groups to certain amino acids in a protein. The most commonly affected amino acids in eukaryotic cells are Ser, Thr and Tyr[246,247]. These residues contain an exposed hydroxyl group cable of being phosphorylated (figure 7.1). Recently the evolution of phosphorylation sites from acidic amino acids has been explored[248]. Protein phosphorylation is a reversible process with many enzymes and receptors in cell signaling pathways being regulated by their state of phosphorylation. Phosphorylation is such a fundamental PTM that two diverse families of enzymes, kinases and phosphatases, have evolved to phosphorylate and dephosphorylate other proteins respectively[249]. Protein phosphorylation, like protein ubiquitination, is responsible for the maintenance of a hugely diverse range of processes in the cell, including the cell cycle, metabolism and neuronal differentiation. It is estimated that approximately 30 % of proteins encoded by the human genome are subject to phosphorylation[246,250]. In fact most proteins are phosphorylated at more than one site in a process known as multisite phosphorylation[251]. The significance of phosphorylation to the world of science was recognized when Edmond Fischer and Edwin Krebs were awarded the Nobel Prize in Physiology and Medicine in 1992 for its discovery[246].



phospho serine



phospho threonine



phospho tyrosine

Figure 7.1 – The structures of the phosphorylated Ser, Thr and Tyr amino acids.

7.1.2 Phosphomimetic mutations in proteins

As discussed, Ser, Thr and Tyr amino acids are subjected to phosphorylation *in vivo*. Mimicking phosphorylation of Ser and Thr amino acids in proteins can be achieved by mutating these residues to Asp or Glu residues. The 3D structures of both Asp and Glu amino acids are very similar to that of phospho Ser and Thr (figure 7.2). Several studies have shown the insertion of such phosphomimetic mutations to proteins which have permitted the study of phosphorylated proteins[252,253,254]. The study of phospho Tyr residues is significantly more complicated. Despite Tyr possessing a more unique structure, substituting for Asp or Glu residues has been shown to be effective in some cases[255,256].

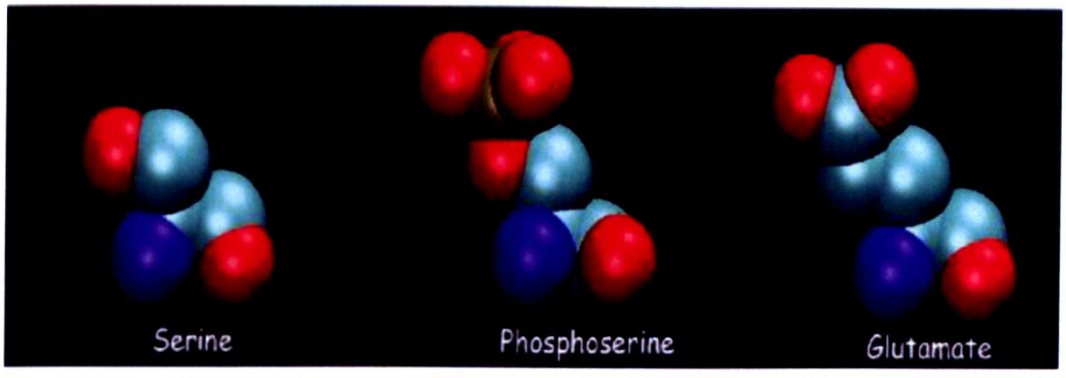


Figure 7.2 – The 3D structures of Ser, phospho-Ser and Glu. The 3D structure of Glu is very similar to phosphor-Ser and is therefore considered an ideal way to mimic phosphorylation *in vitro*.

7.1.3 Ser403 is phosphorylated in the p62 UBA

A recent *in vivo* study by Matsumoto *et al* identified eight phosphorylation sites in the full length p62 protein[257]. Of these eight sites, there were seven phospho Ser and one phospho Thr (figure 7.3). These phosphorylation sites were identified from Neuro2a cells transfected with GFP tagged p62 under proteasome inhibited conditions using liquid chromatography coupled to tandem mass spectrometry. p62 is primarily degraded by autophagy rather than the proteasome, it was therefore necessary to prevent autophagy in order to detect phosphorylated p62. Cells transfected with full length p62 phosphorylated at one of the eight phosphorylation sites were examined. Only Ser 403 phosphorylated p62 levels increased under autophagy inhibited conditions suggesting that Ser 403 phosphorylation may enhance p62 entry to autophagosomes.

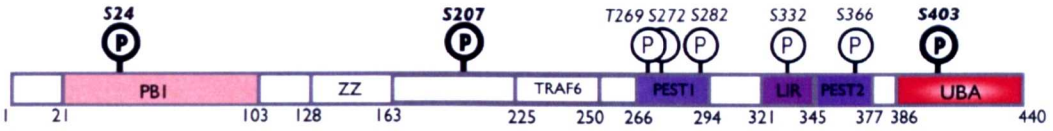


Figure 7.3 – a schematic diagram of the full length p62 protein showing all eight phosphorylation sites. Taken from Matsumoto *et al*[257].

The phosphomimetic mutant S403E was generated in full length p62 and used to investigate the interaction of S403 phosphorylated p62 with ubiquitin. A dephosphorlyated control, S403A, was also engineered in full length p62. The results of the study revealed that the S403E mutant showed an increased affinity for polyubiquitin chains both *in vitro* and *in vivo*. The S403A mutant showed a reduced affinity when compared to wild type p62 UBA. When linkage selectivity was probed by these mutants the S403A mutant displayed a strong preference for Lys63 linked polyubiquitin chains much like wild type p62 UBA. However, the S403E mutant showed enhanced binding for both Lys48 and Lys63 polyubiquitin chains, suggesting that p62 UBA phosphorylation transforms the p62 UBA to a state which is has a higher affinity for ubiquitin.

A protein kinase which phosphorylates the p62 UBA at Ser403, known as CK2, was also identified in the study. The consensus sequence for CK2 is not perfectly matched to the p62 UBA sequence resulting in only partial p62 UBA phosphorylation. The CK2 consensus sequence is S/T-X-X-D/E; whereas, the sequence found in human p62 UBA reads S-M-G-F-S-D. This sequence of the p62 UBA also contains the conserved X-G-F-X motif found in UBA domains which functions to bind to ubiquitin. The motif, Met 404, Gly 405, Phe 406 and Ser 407, as found in human p62 UBA, generates a binding surface that is complementary to ubiquitin, both hydrophobically and electrostatically, facilitating an interaction with the hydrophobic patch centered on Ile 44 in ubiquitin. Ser 403 is located at the C-terminal end of helix 1, immediately prior to the ubiquitin binding motif.

The article concludes that a dynamic equilibrium between the phosphorylated and dephosphorylated form exists *in vivo*. The phosphorylated form becomes favoured when the polyubiquitined protein level is increased permitting the phosho UBA:ubiquitin complex to be recruited to the sequestosome (figure 7.4).

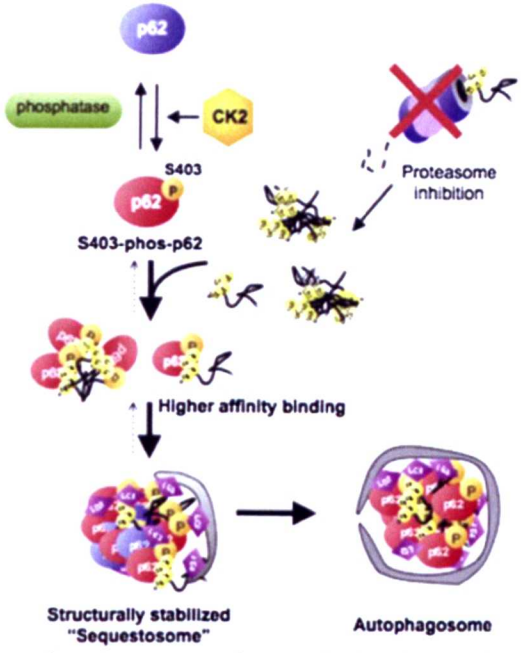


Figure 7.4 – Schematic representation of Ser403 phosphorylated p62 in autophagy. The mechanism was proposed by Matsumoto *et al*[257].

7.1.4 Summary and aims

In full length p62 the phosphomimetic mutant S403E showed enhanced ubiquitin binding relative to wild type p62; whereas, the S403A dephosporylated mutant continued to show weak binding. The main aim of this chapter was to investigate how phosphorylation at Ser403 is able to alter the affinity for ubiquitin using the isolated UBA domain. The p62 UBA mutants S403D and S403E were generated to study the phosphorylated form of the p62 UBA and a control S403A mutant was produced to study the dephosphorylated form.

We aim to probe the structure of phosphorylated p62 UBA and rationalise the basis for ubiquitin recognition using NMR spectroscopy. We also aim to characterise the Ser403 p62 UBA mutants using biophysical techniques to provide an insight into the properties of phosphorylated p62 UBA. Since the previous study only measured affinity using pull down assays, we also aim to calculate affinities for the S403 mutants and compare those to previously obtained values for wild type p62 UBA. Moreover, competitive binding studies using ESI-MS could highlight if a preference for ubiquitin is observed by Ser403 phospho p62 UBA *in vitro*. Biological assays and GST pull down studies, as part of a collaborative project (group of Dr Robert Layfield, University of Nottingham), were also conducted by Miss Alice Goode.

Table 7.1 – A list of the phosphomimetic p62 UBA mutants. The mutated residue is shown for each mutant in red.

Mutant	Helix 1 sequence (residues 392 – 404)
S403A	P R L I E S L S Q M L A M
S403D	P R L I E S L S Q M L D M
S403E	P R L I E S L S Q M L E M

The results of chapter 6 clearly show that weaker p62 UBA dimers have increased affinity for ubiquitin. The equilibrium shifts to favour the biologically active monomer meaning there are more monomeric UBAs available for binding. Another main aim of this chapter was to investigate whether the structural integrity of the UBA dimer had been altered by phosphorylation at Ser403. It was postulated that the equilibrium would be largely unaffected as the position of Ser403 is not found at the dimerisation interface (figure 7.5).

PDB mutations in the p62 UBA have previously been shown to have a slightly reduced affinity for ubiquitin[170, 171]. Such subtle effects are unlikely to dramatically affect downstream signalling in the NF-κB pathway and subsequently cause abnormal osteoclastogenesis. It would be interesting to determine whether phosphorylation in PDB mutants leads to more drastic changes in the affinity the p62 UBA has for ubiquitin. Since the Ser403 phosphorylation site is located close to the ubiquitin binding patch, it is possible that a relationship between ubiquitin binding and phosphorylation exists. Therefore, a combination of the effects of ubiquitin binding, dimerisation and phosphorylation are likely to play a role in the onset of PDB.

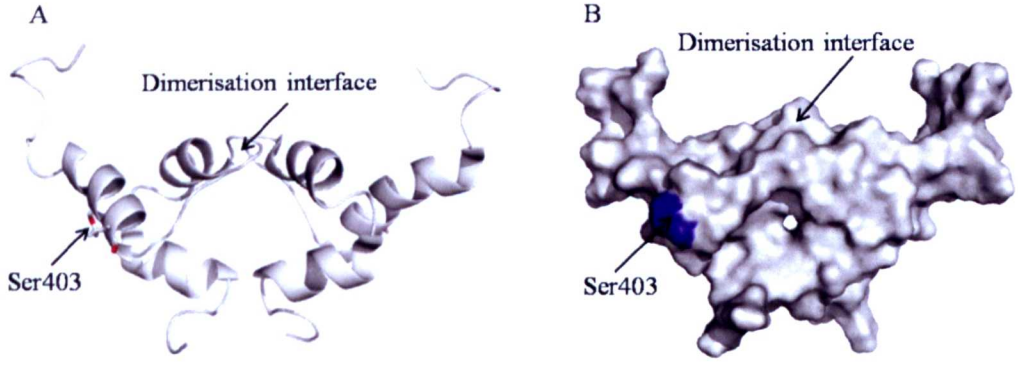


Figure 7. 5 The positioning of Ser403 in the p62 UBA dimer (PDB ID 2KNV). A) the ribbon diagram. B) the surface plot. The C2 symmetry of the p62 UBA dimer prevents both Ser403 residues from being seen.

7.2 Results

7.2.1 Monitoring the monomer-dimer equilibrium in the S403 mutants

7.2.1.1 ESI-MS dilution experiments confirm the presence of a monomer-dimer equilibrium

In order to check how Ser403 affects dimer stability and therefore the monomer-dimer equilibrium, dilution experiments were conducted using ESI-MS. ESI-MS experiments are performed at low sample concentrations which is ideally suited to monitoring the monomer-dimer equilibrium compared to other biophysical techniques. ESI-MS was considered an excellent starting technique to begin studies on the monomer-dimer equilibrium of the various S403 mutants. The same experiments had been used to monitor the equilibrium of the helix 2 dimerisation interface mutants studied in chapter 6.

Initially native ESI-MS was conducted using 5 μM samples to check the molecular masses of all 3 mutants. Samples were prepared by dissolving lyophilized protein in 25 mM ammonium acetate, pH 7 and diluting to the appropriate concentration. A spectrum consisting of multiple charge states was observed in all cases. Mass ions corresponding to the monomeric and dimeric species of each mutant were visualized, but no higher order oligomers. Masses of 5727.0, 5770.9 and 5785.9 Daltons were reported for the S403A, S403D and S403E p62 UBA mutants respectively. The masses recorded by the masslynx software were in excellent agreement with the predicted masses of 5723.4, 5767.4 and 5781.4 Daltons for S403A, S403D and S403E respectively.

Native ESI-MS dilution studies were conducted across the concentration range 8 μM to 1 μM. Initially an 8 μM sample was prepared in 25 mM ammonium acetate, pH 7 and this sample was serially diluted to 1 μM. A spectrum containing multiple charge states was visualized (figure 7.6), as outlined above. Mass ions corresponding to the +2 to +6 charge state of the monomer were observed at 8 μM. The ion corresponding to the +2 charge state was not visible at lower concentrations. The distribution of charge states varied across the concentration range, however the +3 and +4 mass ions tended to be the most abundant in the spectra.

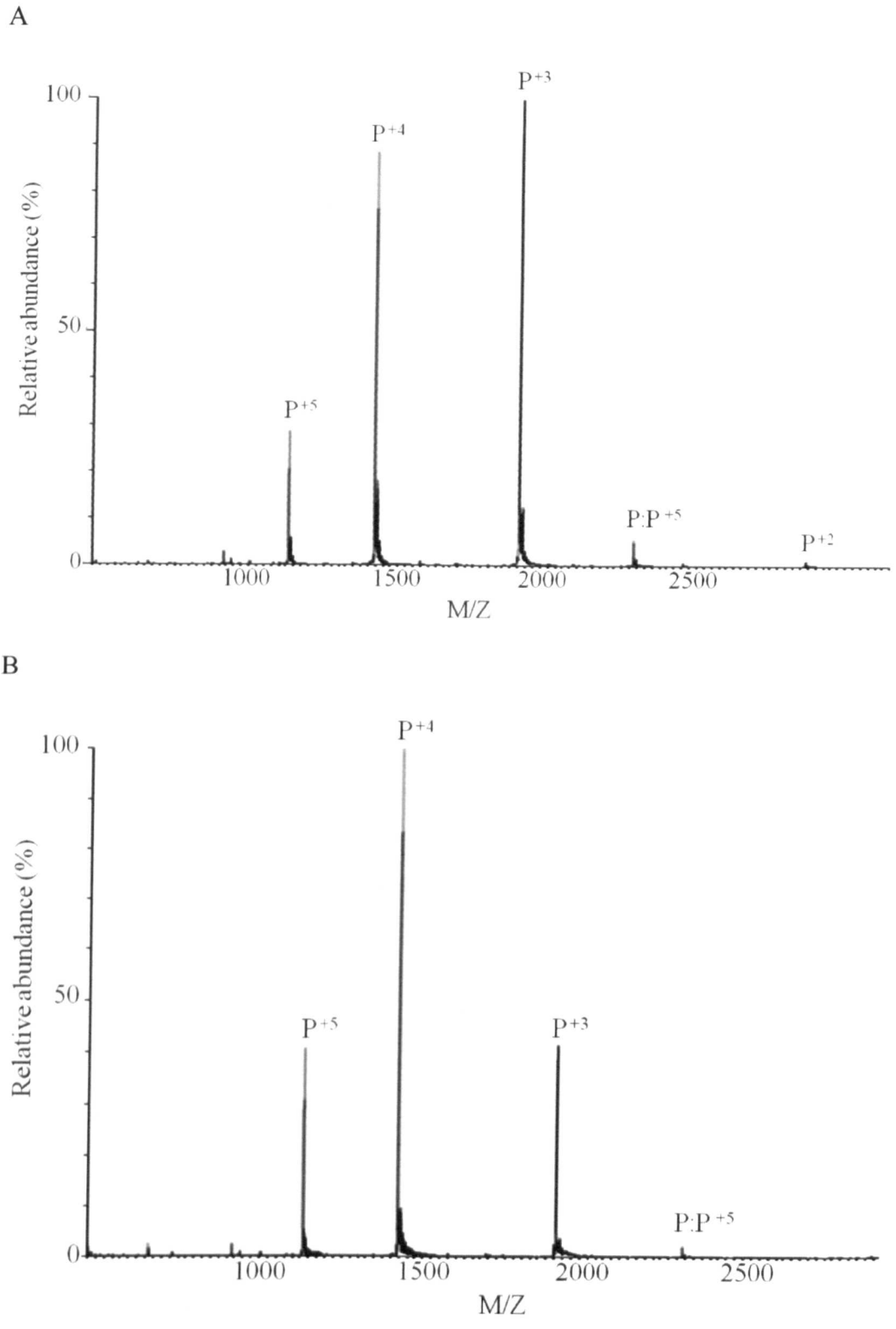


Figure 7.6 – The native ESI-MS spectrum of the S403E p62 UBA mutant at A) 8μM and B) 1μM. Spectra were recorded in 25 mM ammonium acetate, pH 7. The labels correspond to the monomeric species (denoted P) except for the +5 ion for the dimer (denoted P:P) which doesn't overlap with a monomer peak. The changes in the monomer-dimer equilibrium are easily visualized in these two spectra by the changes in intensities by the observed ions.

As discussed in chapter 6, mass ions for the monomer and dimer overlap where both the mass and charge are doubled preventing accurate concentrations of each of the species from being calculated. The m/z value for each charge state recorded different intensities at different concentrations enabling an average charge state to be calculated. Changes in the average charge state have been reported to reflect changes in the solvent exposed surface area[221], which is different for the two different species.

The calculated average charge state was plotted against concentration and compared to values reported for the wild type p62 UBA (figure 7.7). The values recorded for all 3 mutants were similar to the wild type values highlighting minimal changes to the surface exposed surface area, which in theory reflects minimal changes to the monomer-dimer equilibrium. All 3 mutants showed a slightly lower average charge state across the concentration range than those recorded for the wild type. This is indicative of a slight decrease in the solvent exposed surface area for the S403 mutants. The phosphomimetic mutants S403D and S403E have virtually identical average charge states across the concentration range highlighting how these mutants are very similar to one another in terms of solvent exposed surface area throughout the dilution. Moreover, the magnitude of change across the concentration range was comparable to that observed by the wild type p62 UBA for each mutant.

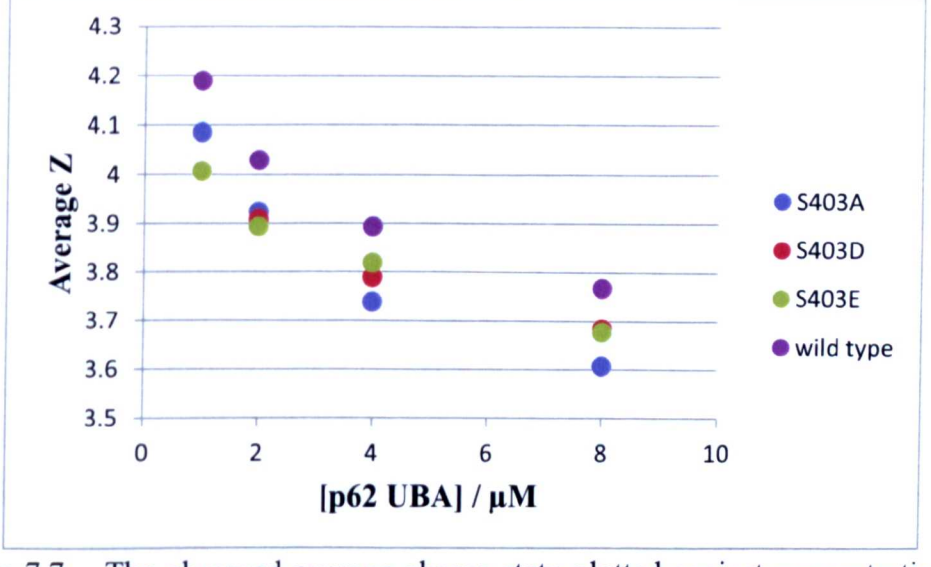


Figure 7.7 – The observed average charge state plotted against concentration of UBA. The wild type data is also shown as a comparison. The data shows that as the concentration decreases the average charge state increases. This change represents the change in populations of monomer and dimer at each concentration.

The native ESI-MS dilution experiments were repeated in the presence of solvent vapour. The technique involved placing a small reservoir of solvent, in this case acetonitrile, into the atmospheric region of the electrospray source. This has the effect of reducing the charge states of the protein ions sprayed from aqueous buffers, such as ammonium acetate[243]. The aim was to see if the intensity of peaks corresponding to the dimeric species would increase in the gas phase. Other biophysical techniques which are conducted in solution, such as NMR and ITC, show the dimer as the dominant species. By decreasing the charge states using solvent exposure, Coulumb repulsion which can drive unfolding and dissociations in the gas phase can be reduced, permitting a more native-like distribution of species.

Using the same samples and conditions as the native ESI-MS dilution experiments, spectra were recorded in the presence of solvent vapour. Approximately 5 ml neat acetonitrile was placed into a small plastic dish in the electrospray source. For all mutants at all concentrations the charge states visualised in the spectra were reduced when compared to the charge states observed in the native ESI-MS spectra. A narrower charge state distribution of +2 to +4 was observed in the presence of the solvent (figure 7.8). In addition to charge state reduction the solvent exposed ESI-MS spectra showed the intensity of the +5 charge state of the dimer increased dramatically. This reflected an increase in dimer population as a result of decreased dimer dissociation. Despite the increase, the monomer remained the predominant species in the gas phase, even at 8 μM. The presence of the solvent to reduce dimer dissociation was not sufficient to cause a completely native-like distribution of species. The ability of the wild type p62 UBA to dissociate so readily in the gas phase is likely to be the result of the predominantly hydrophobic interaction between the two subunits. Hydrophobic interactions have been shown to be significantly weakened in the gas phase[178,179]. The dimerisation interface does not contain the mutation, therefore the hydrophobic contacts between the subunits should not be altered. The S403 dimers are therefore likely to dissociate as readily as the wild type p62 UBA does during ionisation. Since the monomer remained the dominant species in the gas phase under these conditions, all further ESI-MS experiments were completed using standard native conditions.

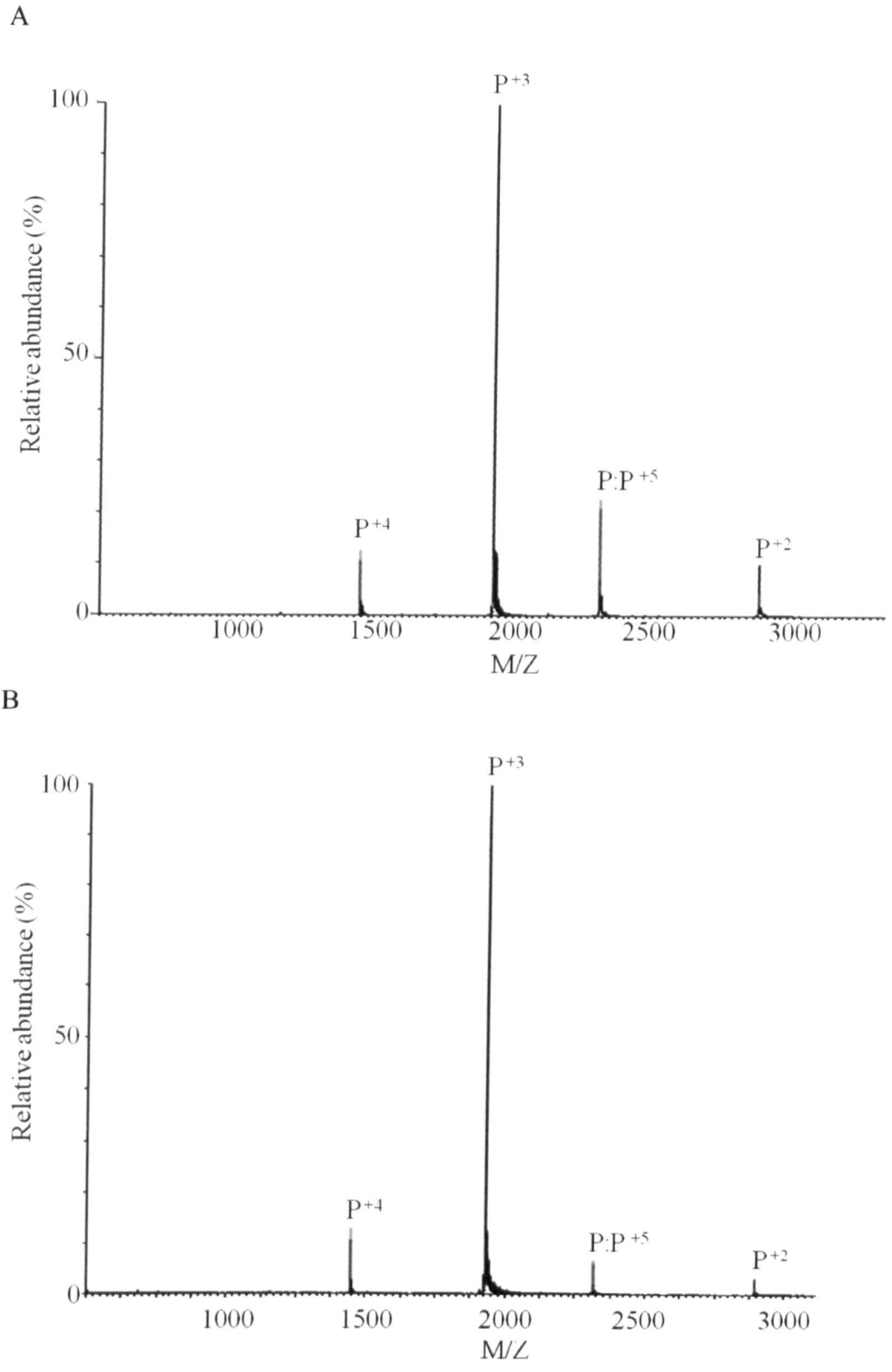


Figure 7.8 – The native ESI-MS spectrum for the S403E p62 UBA mutant in the presence of acetonitrile at A) 8μM and B) 1μM. Spectra were recorded in 25 mM ammonium acetate. A small reservoir of neat acetonitrile was placed in the electrospray source. The labels correspond to the monomeric species (denoted P) except for the +5 ion for the dimer (denoted P:P) which doesn't overlap with a monomer peak. The intensities of the +5 charge state for the dimer have increased in both spectra highlighting more dimer in the gas phase under these conditions.

7.2.1.2 NMR dilution experiments confirm the monomer-dimer equilibrium observed by ESI-MS

In order to reinforce the results of the dilution ESI-MS experiments, dilution experiments were conducted using NMR spectroscopy. Using an 800 MHz spectrometer fitted with a cryoprobe, high quality data using concentrations as low as 10 μM could be obtained. NMR has the advantages that it can provide information on the relative populations of monomer and dimer as well as the extent of protein folding. Experiments using NMR are also completed in solution and are more likely to reflect the *in vivo* environment. Dilution NMR studies were also used to monitor the equilibrium of the helix 2 dimerisation interface mutants studied in chapter 6. The results highlighted that some of the mutations across the helix 2 dimerisation interface caused significant structural changes.

Dilution NMR studies on wild type p62 UBA had previously identified the dimer as the dominant species at concentrations above 6 μM, with 47 resonances corresponding to backbone amide protons observed (see chapter 6 for ^{1}H-^{15}N HSQC spectra). Both the dimer and monomer exhibited a high degree of symmetry, as resonances belonging to each subunit were not observed by NMR. At much lower concentrations the monomer was able to be detected. A 2:1 ratio of dimer:monomer was observed at 10 μM and roughly equal populations were detected at 6 μM.

A ^{1}H-^{15}N HSQC experiment was conducted on a high concentration sample (500 μM) and a low concentration sample (10 μM) of each of the S403 mutants (figure 7.9). Spectra recorded on samples less than 10 μM showed poor resolution, therefore 10 μM was considered to be the lower limit for samples recorded using the 800 MHz spectrometer. Serial dilution experiments were not completed in the interest of saving time. The 500 μM sample was prepared by dissolving lyophilized protein in 25mM potassium phosphate, 25 mM NaCl, pH 7. The 10 μM sample was subsequently made from the 500 μM sample. All spectra were recorded at 298 K.

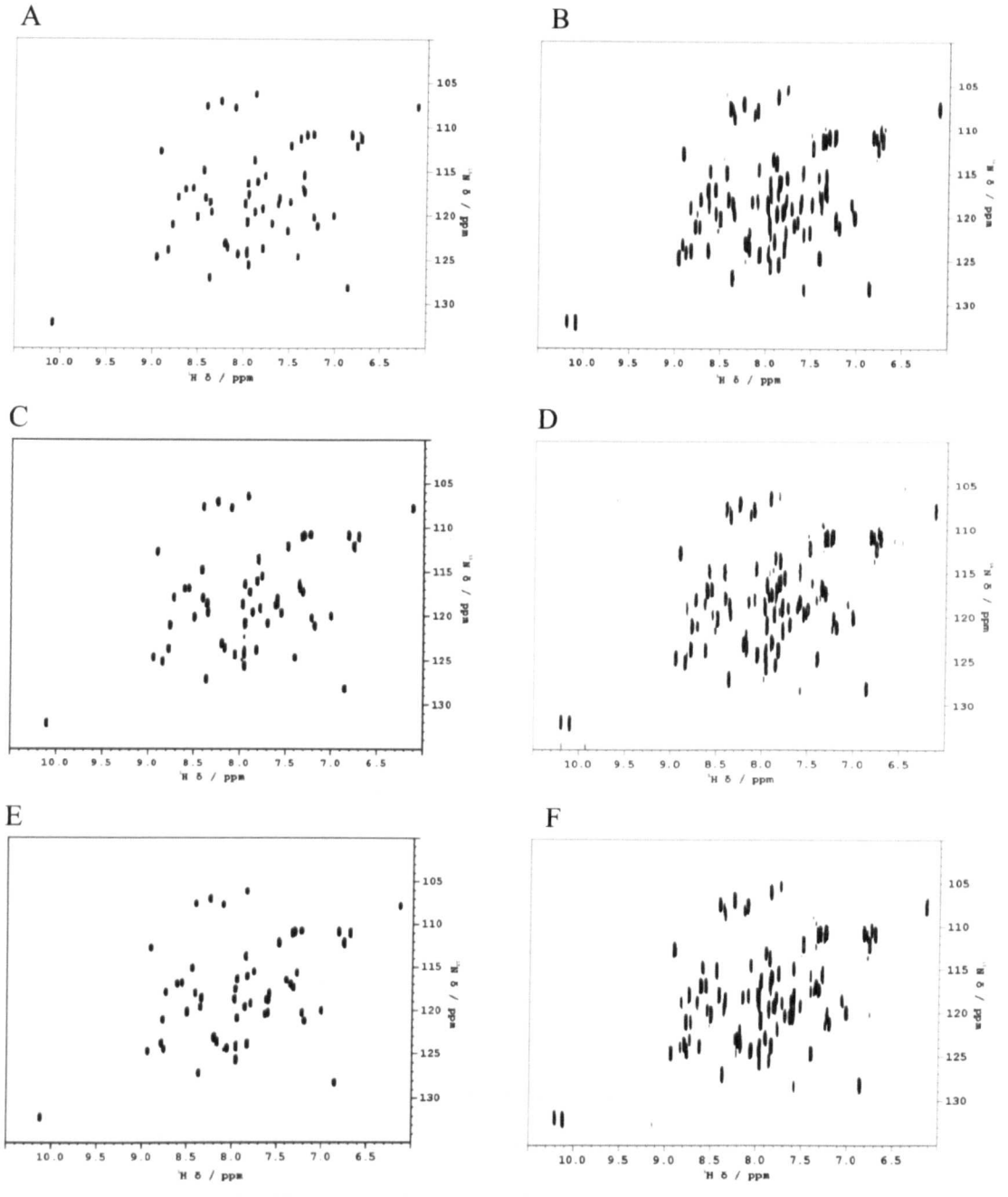


Figure 7.9 - The ^{1}H-^{15}N HSQC spectrum recorded for the S403 p62 UBA mutants: S403A p62 UBA at A) 500 μM and B) 10 μM, S403D p62 UBA at C) 500 μM and D) 10 μM and S403E p62 UBA at E) 500 μM and F) 10 μM. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and at 298 K. At 500 μM resonances which correspond to the dimer only are observed for the S403 mutants. At 10 μM the number of resonances in the spectrum doubles for all 3 mutants. The additional resonances correspond to the monomer. At 10 μM the Trp side chain peaks in the bottom left region of the spectrum reveal a 2:1 dimer:monomer ratio for all 3 S403 mutants. A 2:1 ratio was also observed by wild type p62 UBA at 10 μM. Since the spectra for the S403 mutants at 500 μM and 10 μM is similar to the spectra for the wild type p62 UBA at these concentrations the data suggests that the monomer-dimer equilibrium is in a similar position to wild type p62 UBA.

The results for the three S403 mutants at both concentrations were extremely similar to the results observed for the wild type p62 UBA. At high concentrations 47 peaks corresponding to a single species (the dimer) were observed in the ^{1}H-^{15}N HSQC spectra for the S403 mutants. At low concentration the number of peaks doubles suggesting a concentration dependent change in the observed species. Peaks corresponding to the monomer become visible at lower concentration as the equilibrium shifts to favour the monomer. No substantial differences in peak shape were observed between the two species enabling the intensity of the peak to be used to measure the relative populations of monomer and dimer. By integrating the peaks which correspond to the two Trp side chains in the 10 μM spectrum a ratio of dimer:monomer was calculated for each mutant. At 10 μM all the S403 mutants show a 2: 1ratio, much the same as the results recorded by wild type p62 UBA; highlighting, that the monomer-dimer equilibrium has not been altered by mutation at position 403.

The spectra recorded for all three S403 mutants show a wide dispersion of peaks. For the wild type p62 UBA the distribution of resonances corresponding to each of the species is comparable, but some resonances in helix 3 showed a large change in chemical shift in the monomeric form. The spectra recorded for the S403 mutants show very similar distribution of resonances at both high and low concentrations reflecting that both forms of the mutants have folded into their native conformations. All resonances were observed between 6 and 10.5 ppm, the same spectral width as observed by wild type p62 UBA. Both forms of the mutants are likely to possess native-like tertiary structures. This contrasts the results observed for the T414A and T414K mutants discussed in chapter 6, which showed a narrower dispersion of peaks indicative of the partial unfolding of these mutants.

When the ^{1}H-^{15}N spectrum for the wild type p62 UBA at 500 μM is overlaid with the spectra for three S403 mutants (also 500 μM) it is clear that no major structural rearrangements have occurred. When probed further it was clear that the phosphomimetic mutants caused more chemical shift perturbations than the S403A mutant. The S403A mutant shows perturbations for Gln400 and Met404 only; whereas S403D and S403E show perturbations of more residues which span the mutation site (discussed later in section 7.2.2.1)

7.2.1.3 ITC dilutions confirm that that the structural integrity of the dimer is not significantly altered as similar values of K_{dim} are produced by the S403 mutants

The ESI-MS and NMR dilution studies provided evidence for both a monomeric and dimeric species in equilibrium, with a population distribution very similar to that observed for wild type p62 UBA. In order to obtain some quantitative data to confirm the resulted obtained by NMR, dilution ITC experiments were set up. The apparent *Kd* for the transition from dimer to monomer (K_{dim}) could be calculated using this technique and the data produced compared to values previously reported for wild type p62 UBA.

Samples between 240 μM and 260 μM of the S403A, S403D and S403E were prepared by dissolving lyophilized protein in 50 mM potassium phosphate, 50 mM NaCl, pH 7. The samples were then degassed as required. The concentrated UBA samples were loaded into the syringe (as separate experiments). Sequential injections of UBA into the cell which contained a degassed buffer solution were recorded in order to observe the calorimetric response of dimer dissociation. A series of endothermic heat spikes with decreasing intensity were recorded generating a dimer dissociation curve. The decrease in signal intensity of the endothermic heat spikes correspond to the decrease in dimer dissociation which occurs as the concentration of the UBA in the cell increases. The dissociation curve was fitted to a dimer dissociation model in order to generate a K_{dim} and ΔH from which a full thermodynamic profile can be generated.

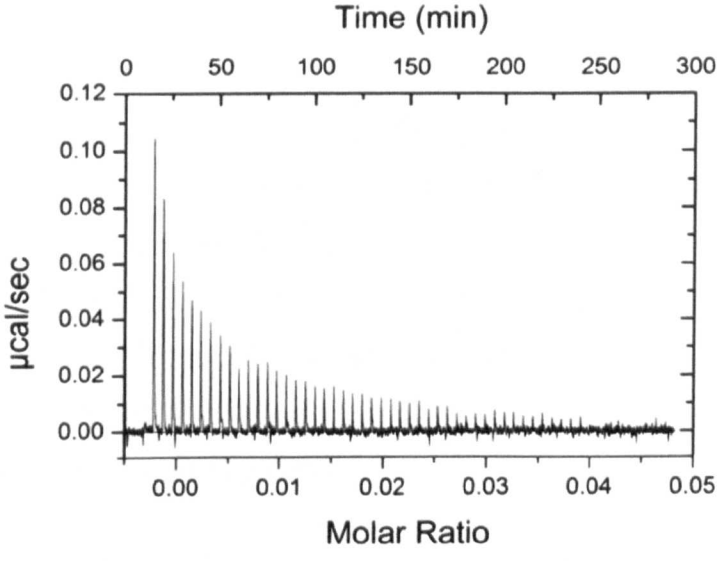


Figure 7.10 – The raw ITC dilution curve for the S403E p62 UBA mutant at 240 μM. This dissociation curve is representative for all three S403 p62 UBA mutants. Experiments were performed in 50 mM potassium phosphate, 50 mM NaCl, pH 7.

All three S403 p62 UBA mutants produced relatively smooth dissociation curves (figure 7.10). The dissociation isotherms for all three mutants fitted well to a dimer dissociation model and produced K_{dim} and ΔH values consistent with data previously reported for wild type p62 UBA. Subsequent calculations of ΔG and ΔS were therefore also in agreement with the thermodynamic profile calculated for wild type protein. K_{dim} values of 3.5 ±0.5 μM, 6.7 ±0.6 μM and 4.2 ±0.5 μM were recorded for the S403A, S403D and S403E mutants respectively (table 6.1). The errors associated with K_{dim} were all small indicating a high degree of accuracy in the fit. The values of K_{dim} for the S403A and S403E mutants were in excellent agreement with 4.1 μM observed for the wild type protein[141]. The S403D mutant recorded the biggest change in K_{dim} when compared to the wild type p62 UBA, however the difference is only a factor of two.

The thermodynamic profiles of the S403 mutants vastly differ to the thermodynamic profile of the weak p62 UBA dimer mutants (T414A, T414K and T419K).

Table 7.1 – the full thermodynamic profiles of the S403 p62 UBA mutants as derived by ITC

Mutant	K_{dim} (±)/ μM	ΔH (±)/ J/mol	ΔS (±)/ J/K/mol	ΔG_{298} (±)/KJ/mol
Wild type	4.1 (±0.6)	60.7 (±1.0)	-102.93 (±1.3)	30.7 (±0.3)
S403A	3.5 (±0.5)	55.4 (±2.2)	-105.5 (±1.1)	31.1 (±0.3)
S403D	6.7(±0.6)	48.8 (±1.1)	-98.9 (±0.7)	29.5 (±0.2)
S403E	4.2(±0.5)	49.5 (±1.7)	-102.8 (±0.9)	30.6 (±0.3)

The ITC data is consistent with both the ESI-MS and NMR data. Overall the data highlights that Ser 403 in wild type p62 UBA is not a critical residue linked to dimer formation or its stability. The positioning of the Ser 403 residue as a residue outside of the dimerisation interface is the likely cause of the consistency in the K_{dim} values recorded for the S403 mutants. Any changes in affinity for ubiquitin by the S403 mutants can therefore not be attributed to a shift in the equilibrium to favour the biologically active monomer.

7.2.2 Investigating how phosphorylation affects binding to ubiquitin using NMR spectroscopy

Phosphorylation at position 403 has been reported to enhance the affinity the p62-UBA has for ubiquitin. In order to investigate how phosphorylation at this site affects the p62-UBA binding, one of the phosphomimetic mutants, namely the S403D p62 UBA mutant was used to conduct forward NMR titrations studies. Unlabelled ubiquitin was therefore titrated into ^{15}N-S403D. Titrations whereby unlabelled ubiquitin was titrated into ^{15}N-S403E and ^{15}N-S403A are being completed by Miss Alice Goode (unpublished data). All NMR experiments were conducted in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and at 298 K. NMR is a technique that can provide information on binding affinity by observing changes in chemical shift to individual residues. The S403 mutants exist in monomer-dimer equilibrium whereby the dimer dissociates to the monomer before the monomer can bind to ubiquitin. In order to visualize the monomer peaks low concentration (10 μM) samples were used for each point within the titration (figure 7.11). The monomer peaks were subsequently used to calculate a K_d for binding by tracking the movement of peaks between the free and bound forms. In order to gain an insight into the residues involved in binding and how that differs to wild type p62 UBA the titration data must be fully assigned.

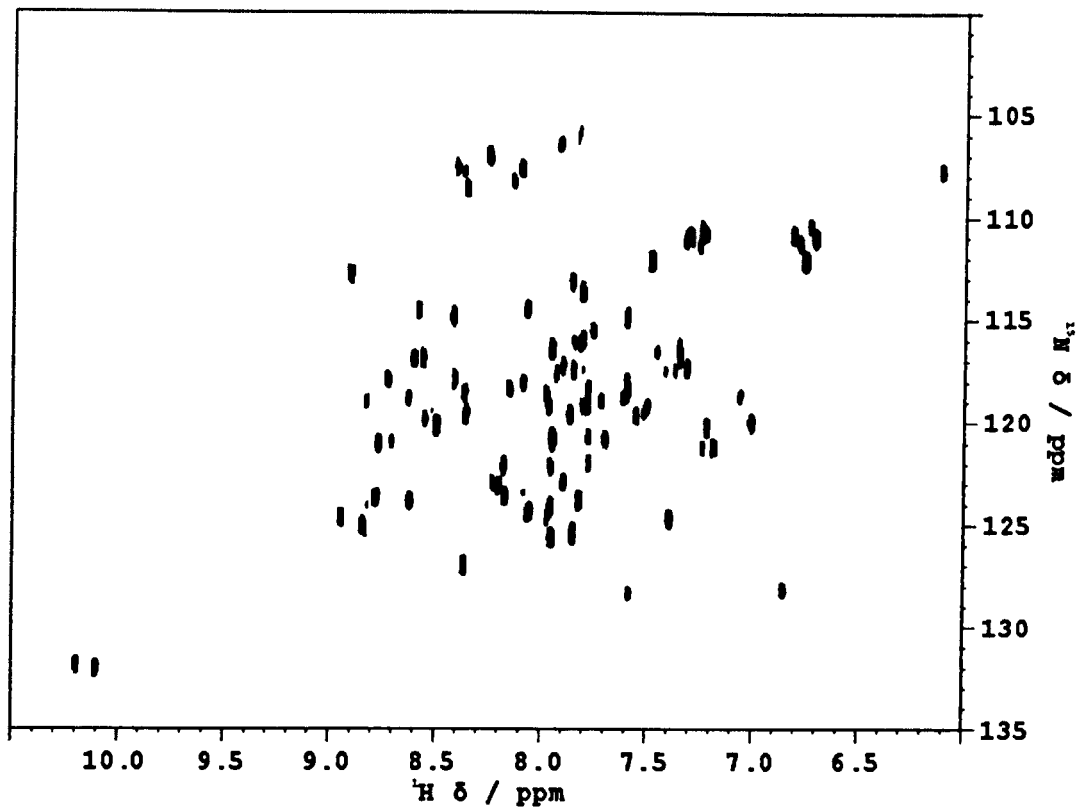


Figure 7.11 – The free ^{15}N-S403D p62 UBA ^{1}H-^{15}N HSQC spectrum at 10 μM. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K. The spectrum shows double the number of peaks compared to a spectrum at 500 μM due to the presence of both the monomer and dimer species.

7.2.2.1 Assigning the free dimer and bound monomer of S403D

The 10 μM sample of the free ^{15}N-S403D p62 UBA showed resonances corresponding to the monomeric and dimeric species (figure 7.11). Of these resonances 39 were assigned to the monomer and 47 to the dimer. Most of the resonances corresponding to both species were easily assigned as they were located at the same chemical shifts as those for the wild type p62 UBA. It was easiest to transfer assignments from the wild type dimer to the ^{15}N-S403D dimer and then transfer the assignments to the 10 μM ^{15}N –S403D spectrum. The 10 μM spectrum was significantly more complicated as it contained almost double the number of resonances. The monomer peaks were assigned once the dimer peaks had been assigned, against based on chemical shifts reported for the wild type monomer.

A spectrum of the ^{15}N-S403D spectrum at 500 μM, whereby only the dimer peaks were visible, showed most resonances were located at the same chemical shift as reported for wild type p62 UBA. Significant differences in chemical shifts were observed for a small number of resonances (figure 7.12 A). It was therefore assumed the resonances corresponding to the same residues in the free monomer would also show differences in chemical shift in the 10 μM spectrum. These resonances corresponded to residues Gln400, Met401, Leu402, Asp403, Met404, Gly405 and Phe406 in the wild type protein. These changes in chemical shift are for residues located in close proximity to the mutated residue Asp403 only.

The fully bound ^{15}N-S403D spectrum showed 46 assignable backbone resonances corresponding to the bound monomer form (figure 7.12 B). There was a strong correlation between the wild type and the ^{15}N-S403D spectra, however differences in chemical shift were observed again for some resonances. The resonances which had moved corresponded to the same residues which had moved in the free form. These residues, Gln400, Met401, Leu402, Asp403, Met404, Gly405 and Phe406, spanned the mutation site. It is noticeable that the MGF motif, residues Met404, Gly405 and Phe406, are perturbed in the ^{15}N-S403D bound spectrum highlighting that these residues have changed their chemical environment and therefore their position in this complex, perhaps facilitating a stronger interaction with ubiquitin.

Of the residues which showed large chemical shift deviations from wild type p62 UBA Leu402 and Gly405 were still able to be assigned based on the wild type assignments. However further experiments needed to be completed to assign residues Gln400, Met401, Asp403 (the mutated residue), Met404 and Phe406 in both the free monomer and dimer. A NOESY and TOCSY experiment was run on a 500 μM sample to complete the assignment of the dimer. In addition to this, a ^{15}N-HSQC-NOESY and ^{15}N-HSQC-TOCSY was run on ^{15}N-S403D fully bound sample (1:4 ^{15}N-S403D: unlabelled ubiquitin) to complete the assignment for the bound monomer. The type of amino acid could be identified using the TOCSY and ^{15}N-HSQC TOCSY for the dimer and bound monomer respectively. Each amino acid has a distinct chemical shift correlation pattern to its backbone NH. Sequential assignment of the cross peaks of the NH region were used in the NOESY and ^{15}N-HSQC NOESY (again for the dimer and bound monomer respectively) where the amino acid type could not be easily identified.

Once the bound monomer peaks were assigned the peaks could be tracked back to the free form as the binding event occurs in fast exchange. For the wild type protein at low concentration (10 μM), resonances for monomeric residues Glu389, Arg393, Met401, Leu402 and Glu409 were visible but these residues were not assignable in the 10 μM ^{15}N-S403D spectrum. A 10 μM NOESY to aid assignment was recorded but the data collected was of poor quality and was unable to aid the assignment of the remaining residues. The peak for dimeric Gly410 is very weak and not visible as a monomer in the wild type spectrum. The peak for Gly410 is undetectable as either a dimer or a monomer for the ^{15}N-S403D. The monomer peaks which were not assigned could potentially have overlapping chemical shifts with the dimer. These would therefore need to be distinguished from one another using 3D NMR techniques. Since nearly all resonances were able to be assigned in the 10 μM ^{15}N-S403D spectrum it was decided that 3D data collection was unnecessary.

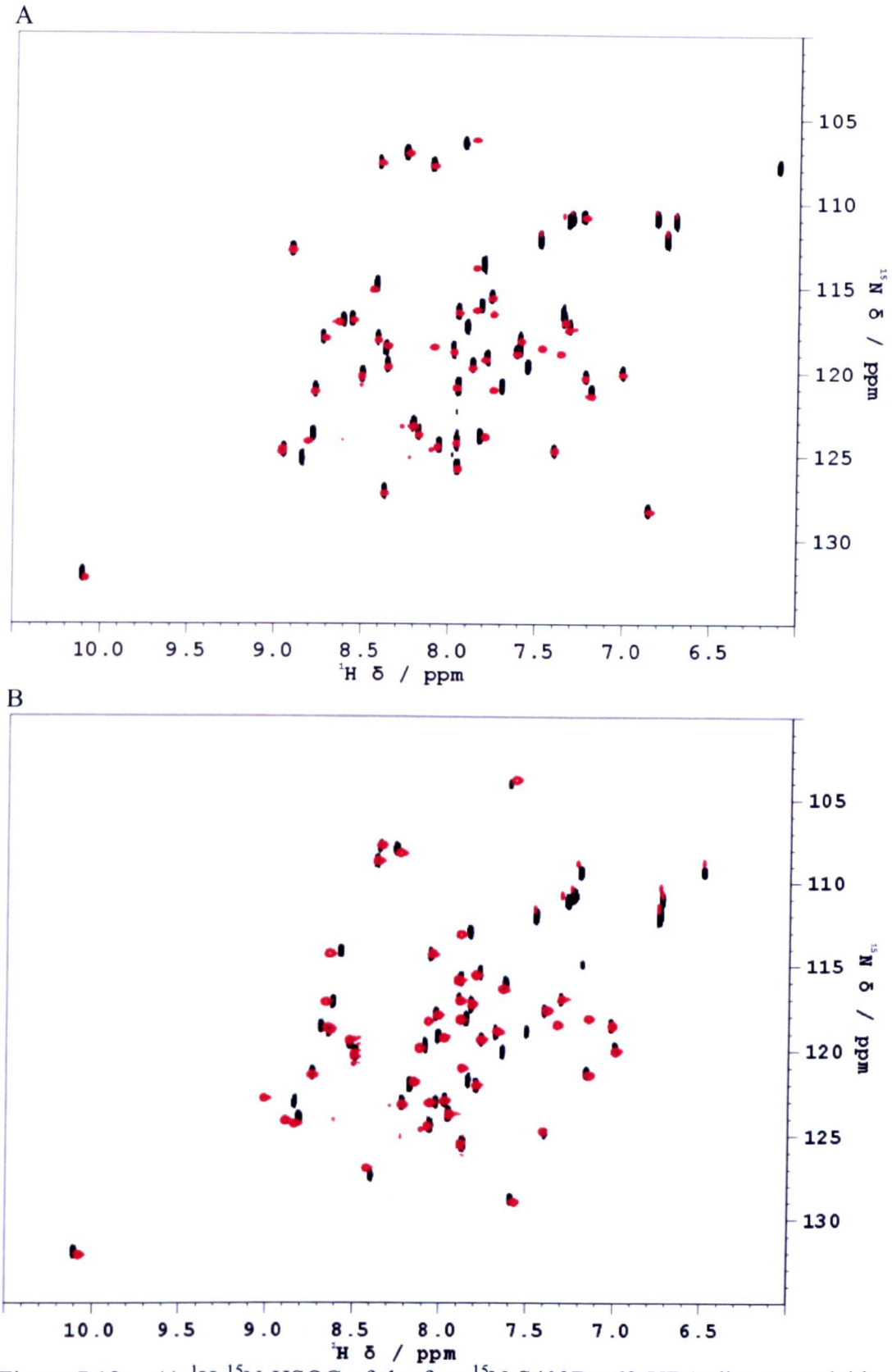


Figure 7.12 – A) ^{1}H-^{15}N HSQC of the free ^{15}N-S403D p62 UBA dimer overlaid with the free wild type p62 UBA dimer (black and red respectively). B) ^{1}H-^{15}N HSQC of the ^{15}N-S403D p62 UBA bound monomer overlaid with the wild type p62 UBA bound monomer spectrum (black and red respectively). Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH7 and 298 K.

7.2.2.2 Identifying which residues on the S403D mutant are involved in ubiquitin binding (forward titration using NMR spectroscopy)

Monitoring the changes in chemical shift for the monomer resonances between the free and bound forms will allow a K_d for binding to be calculated. The binding event occurs as a fast event on the NMR timescale meaning that the peaks which are visualized are a population weighted average of the free and bound forms. Peak movement is therefore easy to track and the CSP can be used to calculate the affinity the S403D mutant has for ubiquitin. Using a 10 μM ^{15}N-S403D p62 UBA sample, peaks for both species are observed in a ratio of 2:1 dimer:monomer. The low signal intensities recorded at 10 μM precluded the use of lower concentrations for titration purposes.

The wild type monomer has been reported to be unstable[141]. This instability was also observed for the S403D mutant. An initial attempt to complete a titration using the methodology utilised by other titrations in this thesis resulted in the sample undergoing sufficient proteolysis causing loss of NMR signal mid way through data collection. A new strategy involving fresh 10 μM ^{15}N-S403D samples combined with lyophilized ubiquitin of known concentrations for each point was adopted.

Up to 20 molar equivalents (i.e 200 μM) of ubiquitin was titrated into 10 μM ^{15}N-S403D (figure 7.13). The results showed concentration dependent changes for both the monomeric and dimeric species. Saturation occurred after 10 molar equivalents (100 μM) of ubiquitin had been added. Saturation for the wild type p62 UBA occurred at the addition of 20 molar equivalents (200 μM) of ubiquitin, highlighting a greater affinity of the ^{15}N-S403D mutant for ubiquitin. As the ubiquitin is titrated in the monomer-dimer equilibrium is shifted to favour the monomer. Similar to the transition from dimer to monomer for the wild type p62 UBA, this event occurs in slow exchange. Consequently, the peaks corresponding to the ^{15}N-S403D dimer decrease in intensity simultaneous to an increase in intensity for the monomer peaks. The peak intensities are therefore proportional to the populations of each species. Resonances belonging to the dimer showed a significant decrease in peak intensity in the presence of 2 molar equivalents (20 μM) of ubiquitin. Eventually the peaks belonging to the dimer disappeared

completely, leaving only the monomer peaks, which are capable of binding to ubiquitin.

Peaks corresponding to the free ^{15}N-S403D monomer move in a concentration dependent manner to the bound form. Fast exchange is seen for nearly all resonances corresponding to the monomeric species. Substantial peak broadening occurs for peaks corresponding to Gly405, Gly425 and Leu428 during the titration. However, these residues are in intermediate exchange and thus return to measurable intensity before saturation occurs.

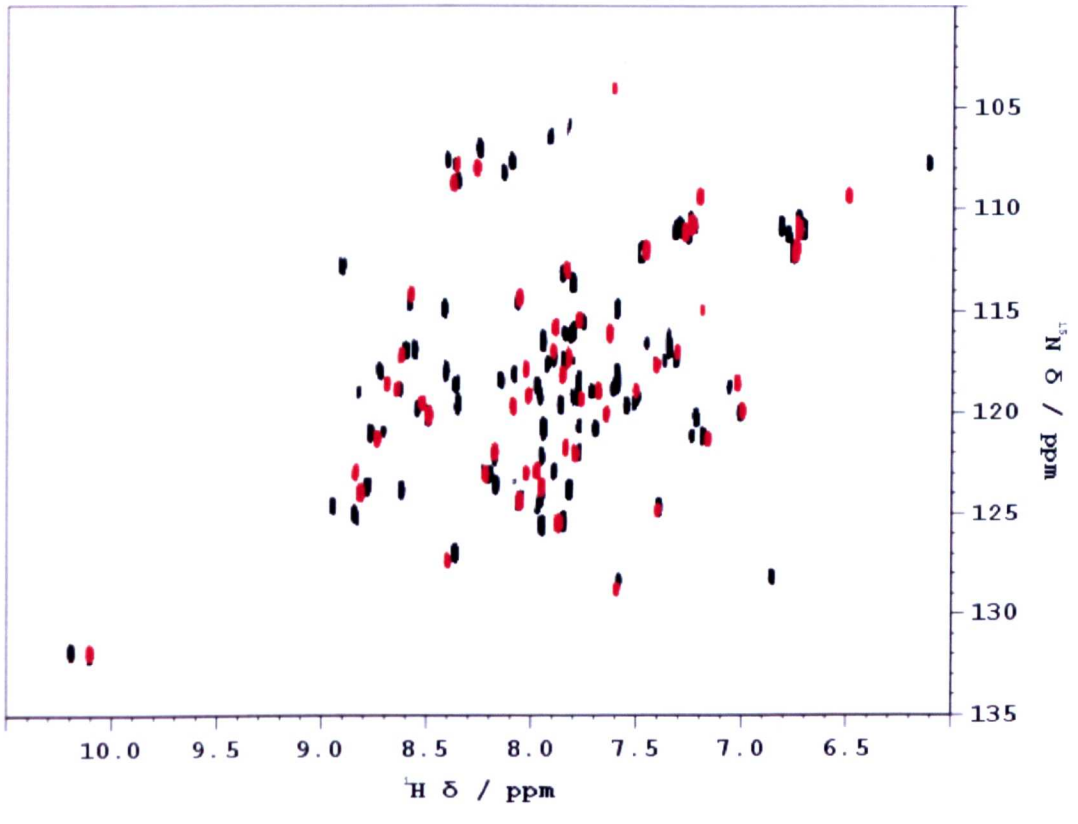


Figure 7.13 – ^{1}H-^{15}N HSQC spectra for ^{15}N-S403D p62 UBA at 10 μM showing peaks corresponding to the monomer and dimer (black) overlaid with the ^{15}N-S403D saturated with ubiquitin (red). The spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K.

The CSPs for the monomer (39 peak pairs correlating to the free and bound forms) were plotted against the residue number to produce a residue specific CSP plot (figure 7.14). The gaps in the plot are where a free and bound peak could not be assigned for that residue. The pattern of CSPs was highly similar to that observed for the wild type p62 UBA monomer, with the biggest CSPs being found

at the end of helix 1 and loop 1. Moreover larger CSPs were observed for helix 3, again correlating well to the wild type data. The CSP plots clearly show the ^{15}N-S403D mutant still utilizes the conserved MGF motif in loop 1 to bind to ubiquitin.

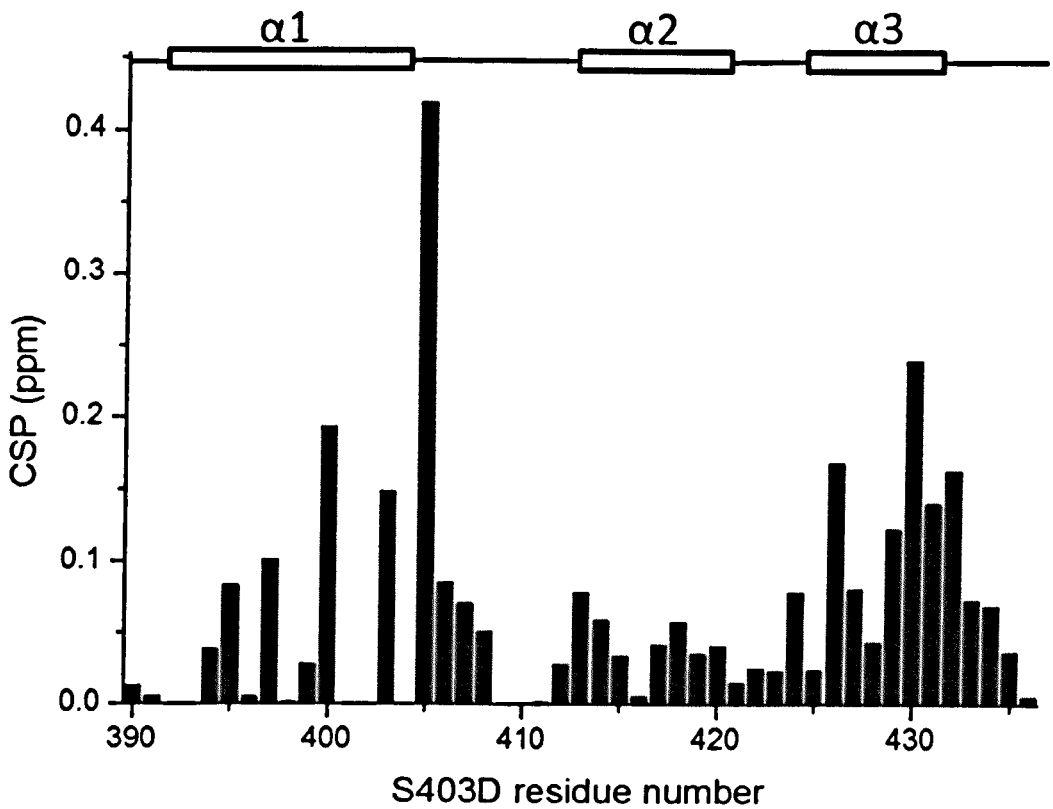


Figure 7.14 –Residue specific CSP for the backbone NH groups for the ^{15}N-S403D p62 UBA monomer between the free and bound forms. The gaps correspond to residues which were unable to be assigned. The largest CSPS are seen for Gly405 in the conserved MGF motif in loop 1 and residues in helix 3. These large CSPs highlight the S403D mutant binds to ubiquitin using a similar mechanism to wild type p62 UBA.

The residues which showed a CSP of more than 0.05 ppm were considered to be significant movers. These residues were plotted onto the surface of the wild type bound monomeric UBA to identify the binding surface (figure 7.15, PDB ID 2JY8). The darker the red colour, the greater the change in chemical shift. Peaks for the monomeric Leu402 and Met404 were not able to be detected and assigned. If these resides were assigned, it was postulated that they would have large CSPs and a clear binding patch centered on the MGF motif would be identified. The MGF loop appears to be utilized by the ^{15}N-S0403D mutant meaning an interaction with the canonical Ile44/Val70 patch on ubiquitin is likely to occur.

The MGF motif forms a surface which protrudes from the 3D structure of the p62 UBA monomer, which is capable of binding to ubiquitin. The observed large CSPs for the MGF loop between the wild type and the ^{15}N-S403D mutant suggests that these residues have changed their orientation to favour a tighter affinity interaction.

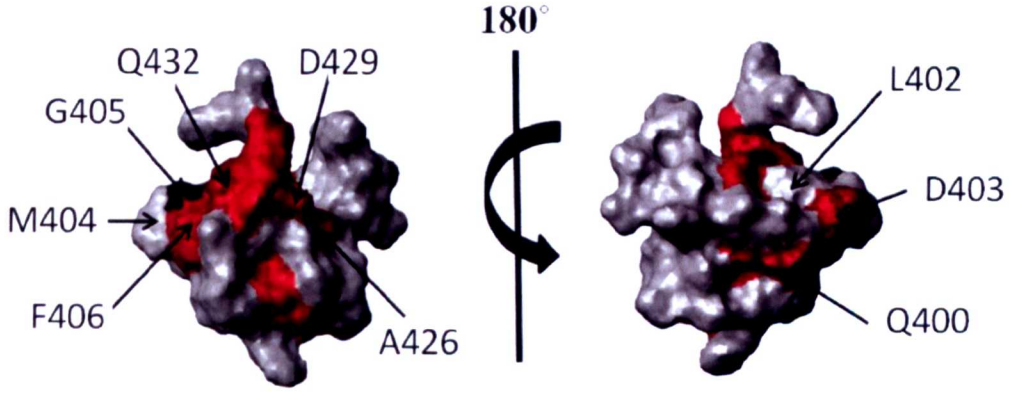


Figure 7.15 – CSPs greater than 0.05 for the ^{15}N-S403D mutant plotted onto the electrostatic surface of the bound monomeric UBA (PDB ID 2JY8). The darker the red colour, the greater the perturbation. The observed CSPs are caused by ubiquitin binding. Residues in loop 1 and helix 3 have the greatest CSPs, highlighting that the S403D binds to ubiquitin using a similar mechanism to wild type p62 UBA.

By globally fitting the residues with large changes in chemical shift between the free and bound forms (CSP >0.05 ppm) to a simple 1:1 binding model using IGOR Pro an apparent K_d could be produced. A total of 19 curves were fitted to generate a value of 229 ±34 μM. The curves fitted well to a 1:1 binding model. The K_d value is best described as the observed K_d (K_{obs}) because the fit does not take into consideration the competitive processes of dimerisation and binding. By using the relationship between K_{obs} and K_{dim} (equation 12, chapter 3), which takes into account the two competing equilibria, a more accurate K_d value can be determined. The S403D mutant was shown to have a K_{dim} of 6.7 ± 0.6 μM as calculated using ITC dilutions (section 7.2.1.3). Therefore a K_d of 39.1 ± 4.7μM was calculated for the interaction between the S403D monomer and ubiquitin. The wild type p62 UBA monomer had been previously shown to have a K_d of 40 ± 10 μM[141]. Rather unusually the S403D mutant therefore does not show a significant increase in affinity relative to the wild type. However, small changes in affinity for ubiquitin are also observed by many of the PDB mutations.

7.2.2.3 Identifying the binding site for the phosphomimetic mutants on ubiquitin (reverse titrations using NMR)

In order to investigate how the phosphomimetic mutants affect binding to ubiquitin unlabelled S403D and S403E p62 UBA were titrated into ^{15}N-ubiquitin. The dephoshorylated mutant S403A was also titrated into ^{15}N-ubiquitin as a control. The wild type p62-UBA has been shown to bind to the canonical Ile44/Val70 binding site. Since NMR spectroscopy can provide information on each residue involved in binding it can identify if the canonical ubiquitin binding site is still utilized by these mutants. Assignments for wild type ubiquitin have been previously determined[8] and could therefore be used to assign the data obtained from the reverse titrations.

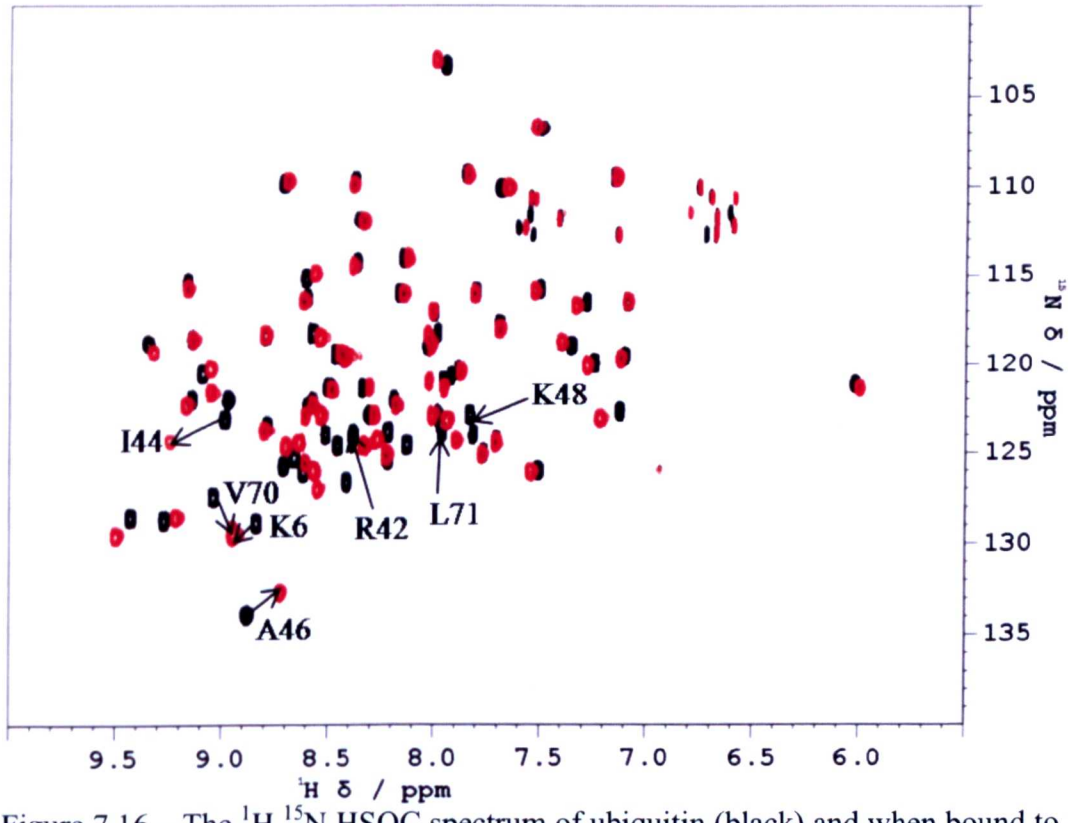


Figure 7.16 – The ^{1}H ^{15}N HSQC spectrum of ubiquitin (black) and when bound to S403E p62 UBA mutant (red). The spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K. The I44 and V70 residues which form the canonical ubiqutin binding patch are highlighted as well as other residues which move significantly (A46 and K6). Residues R42, K48 and L71 which broaden upon titration with mutant UBA and do not reappear are also labelled. This spectrum is representative for all three S403 p62 UBA mutants.

The addition of up to 2 mM mutant UBA was titrated in to 500 μM ^{15}N-ubiquitin. Concentration dependent chemical shifts were observed for each residue consistent with a binding event (figure 7.16). Saturation was observed for the phosphomimetic mutants at 500 μM (a 1:1 mutant UBA : ubiquitin molar ratio). The S403A control mutant was saturated at 1 mM, a 1:2 molar ratio, indicating that this mutant has weaker binding to ubiquitin when compared to the phosphomimetic mutants. ^{15}N-ubiquitin is saturated by wild type p62 UBA after the addition of 4 molar ratios of wild type p62 UBA.

The binding event is fast on the NMR timescale resulting in peaks which are in fast exchange. The only exceptions to this are peaks corresponding to residues Arg42, Ile44, Lys48, Gly53 and Leu71 which are all in intermediate exchange. In intermediate exchange the transition between the signals observed for fast and slow exchange is seen causing broad and often non-observable signals. Peaks for Arg42, Lys48 and Leu71 become undetectable after 100 μM unlabelled mutant UBA has been added and do not reappear in the concentration range used in these experiments. These peaks are all located within the canonical Ile44/Val70 binding patch, which is not unusual as intermediate exchange is often observed for residues found in binding patches. The peaks for Arg42 and Leu71 have weak intensities and both partially overlay with other peaks in the free ubiquitin spectrum. The initial weak intensities of these signals make it more likely that the signal will be lost during the titration. The peaks corresponding to Ile44 and Gly53 are also in intermediate exchange with signals that broaden at 100 μM and 250 μM respectively. These peaks, unlike peaks for Arg42, Lys48 and Leu71, sharpen to strong detectable signals when nearing the saturation point. Typically peaks which are in intermediate exchange would be expected to sharpen to considerable intensity once the fully saturated form is reached, however this does not seem to be the case for residues Arg42, Lys48 and Leu71. These particular residues have been shown to behave in the same manner when wild type p62-UBA is titrated into ^{15}N-ubiquitin, highlighting a conserved mechanism of binding between the wild type protein and the S403 mutants.

As previously mentioned, the vast majority of peaks are in fast exchange during the binding event. This means that the observed chemical shift at any concentration is a weighted average of the chemical shifts observed for the free

and fully bound forms of the protein. The CSP between the free and bound forms can be plotted against residue number to observe changes on a per residue basis (figure 7.17). The greatest CSPs in all cases were observed for the Ile44 and Val70 residues confirming that the canonical ubiquitin binding site is utilized by these mutants. The movement of the Val70 peak in the titration with the S403A mutant was not as great as it was for either of the phosphomimetic mutants. In the titrations with S403D and S403E p62 UBA, the Val70 peak merges with the peak for Lys6; however, in the S403A titration there are two distinct peaks close in proximity but not merged. Two separate peaks for Val70 and Lys6 are also observed for ubiquitin when wild type p62 UBA is titrated in. The CSP for Val70 is greater for the S403A mutant than it is for wild type protein indicating a stronger interaction for the S403A mutant and ubiquitin when compared to wild type p62 UBA. The Val70 CSP is much higher for the S403D and S403E mutant again highlighting a much tighter interaction with ubiquitin than observed for the wild type p62-UBA or S403A mutant.

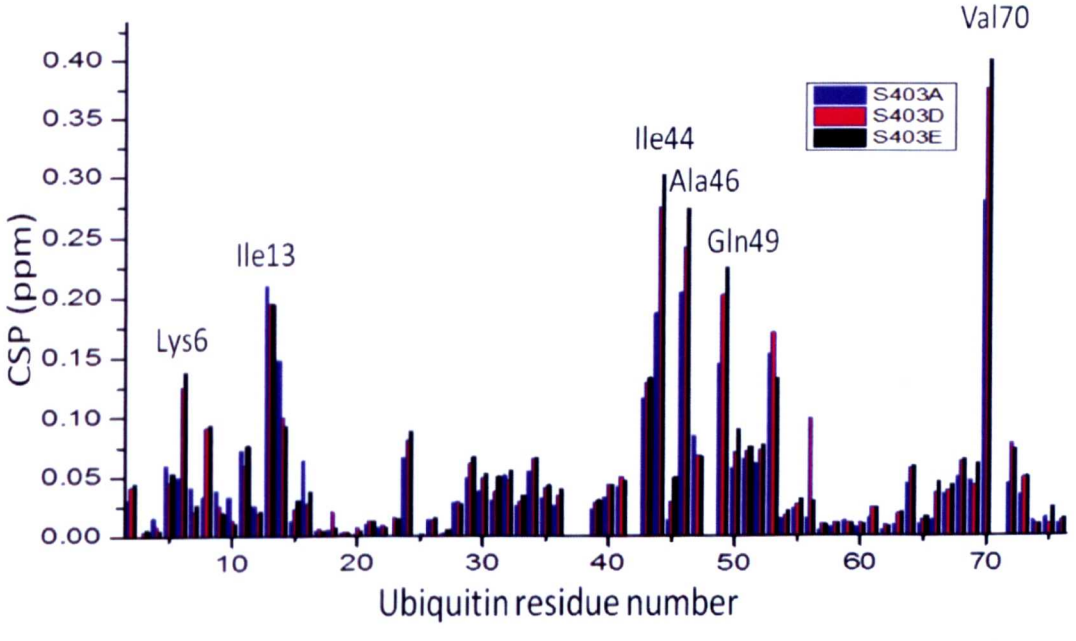


Figure 7.17 – Residue specific CSPs for ubiquitin when bound to the different S403 p62 UBA mutants: S403A (blue), S403D (red) and S403E (black). The CSPs were generated from reverse NMR titrations whereby unlabelled p62 UBA mutants were titrated into ^{15}N-ubiquitin. The strong correlation in the data highlights that all 3 mutants bind the canonical Ile44/Val70 binding patch on ubiquitin.

When you compare at the CSPs for the S403D and S403E mutants with the S403A value you can determine the effects of mimicking phosphorylation compared to the control on a per residue basis (figure 7.18). Two patches of perturbed residues are observed which centre on the Ile44 and Val70 residues. In addition to this Lys6 shows a large CSP as a result of phosphorylation.

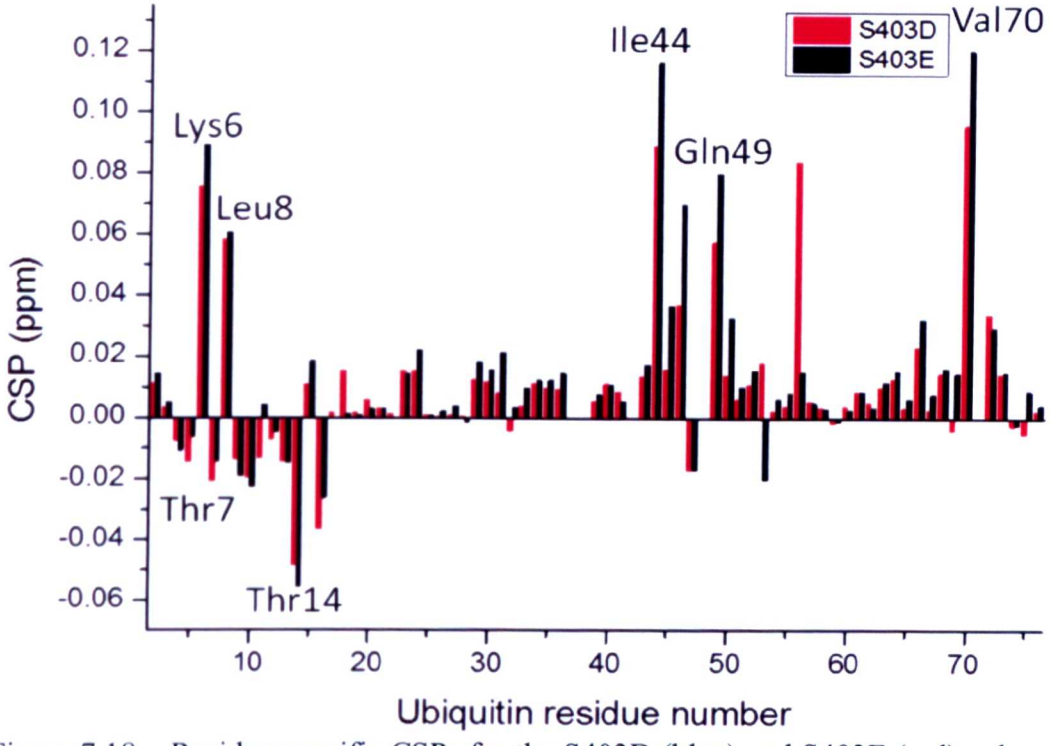


Figure 7.18 – Residue specific CSPs for the S403D (blue) and S403E (red) values minus the S403A CSP values. The effects of phosphorylation are visualised per residue. The residues most affected by phosphorylation are Lys6, Leu8, Ile44 and Val70.

Residues with large CSPs were plotted onto the surface of ubiquitin (figure 7.19, PDB ID 1UBQ). A large patch centered on the cannonical Ile44/Val70 hydrohobic patch, confirming that this is indeed the binding patch utilized in binding to the S403 mutants. The darker red colour indicates the residues which are more perturbed by binding. The black residues correspond to Arg42, Lys48 and Leu71, which are in intermediate exchange. From these surface plots it is clear that the three S403 mutants are affected in slightly different ways. The phosphomimetic mutants S403D and S403E have more residues which are affected by binding.

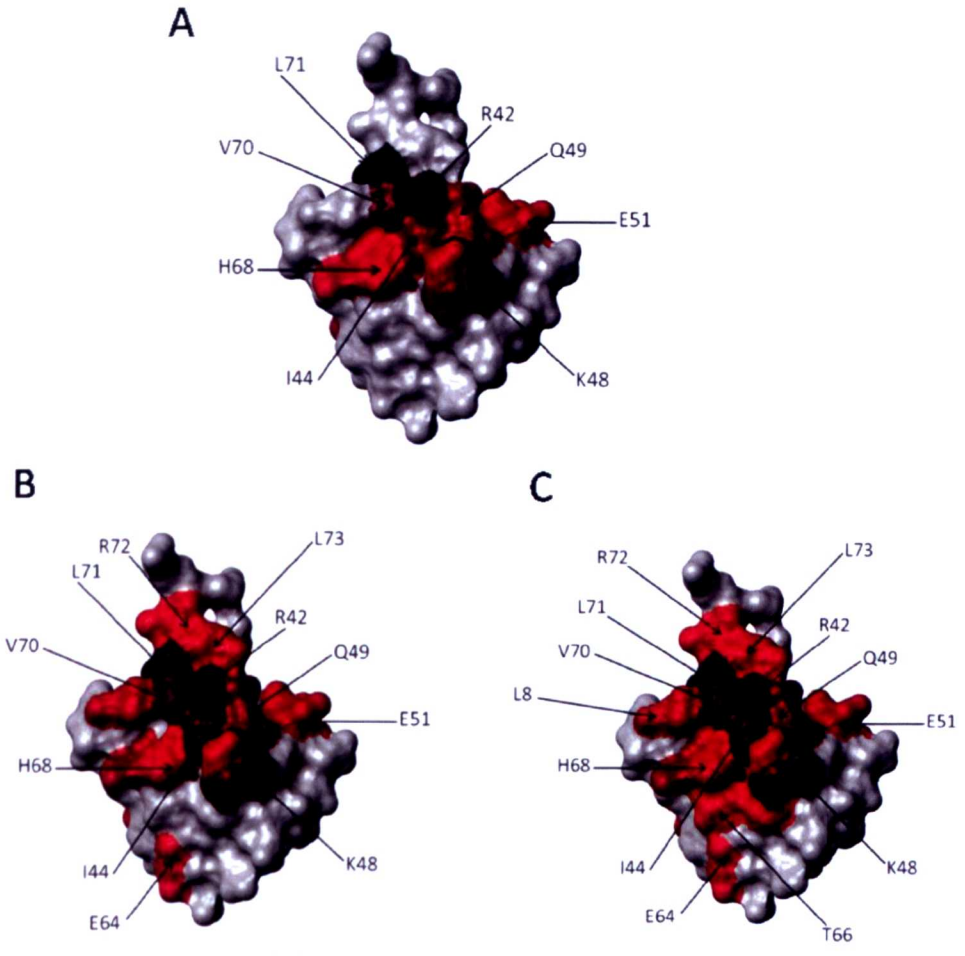


Figure 7.19 – CSPs used for binding to the S403 mutant UBAs plotted onto the surface of ubiquitin (PDB ID 1UBQ): A) S403A p62 UBA, B) S403D p62 UBA and C) S403E p62 UBA. The darker the red, the larger the CSP. Residues Arg42, Lys48 and Leu7 broaden during the titrations and are coloured black. A more extensive binding patch is revealed in the phosphomimtetic mutants indicative of tighter binding interface.

The residues which had moved significantly (CSP > 0.05 ppm) between the free and bound forms were used to calculate the binding affinity (K_d) of a the interaction. K_d values of 373 ± 19 μM, 337 ± 11μM and 319 ± 11μM were calculated for the S403A, S403D and S403E mutants respectively. These values were obtained by globally fitting all residues with a CSP greater than 0.05 ppm to a 1:1 binding model in IGOR Pro. The full details are listed in table 6.2. A value of 540 ± 45 μM was previously reported for wild type p62 UBA[140]. The K_d values calculated for all the S403 mutants are lower indicating that all three mutants have a stronger interaction with ubiquitin than wild type p62 UBA. Despite there being only subtle differences in K_d, there is a slight preference for the phosphomimetic S403D and S403E mutants over the S403A non-phosphorylated control.

Table 7.2 – The K_d values produced from globally fitting residues with a CSP greater than 0.05 ppm as calculated by IGOR Pro.

p62 UBA mutant	K_d (μM)	Number of residues with a CSP >0.05 ppm	Saturation point (Ub;UBA molar ratio)
Wild type	540 (±45)	12	4:1
S403A	373 (±19)	18	2:1
S403D	337 (±11)	22	1:1
S403E	319 (±11)	26	1:1

Although the reverse titrations have the advantage of ubiquitin moving from its free to bound forms using fast exchange, it also has the disadvantage that the p62 UBA is the invisible part of the NMR experiment. This therefore means that the K_d values calculated from global fitting are not a true reflection of binding alone, since dimerisation of the UBA and ubiquitin binding are competitive processes[141]. The dissociation of the dimer to the monomer which precedes binding is not taken into consideration by IGOR Pro as the determined values are the product of the two competing equilibria. Hence the previously determined K_d is best described as the observed K_d (K_{obs}). Values of K_{dim} were calculated using ITC dilutions (section 7.2.1.3) and by using the relationship between K_{obs} and K_{dim} (equation 12, chapter 3), which takes into account the competitive processes of dimerisation and binding in the case of the p62 UBA, a more accurate K_d value to be determined. Based on the values of K_{obs} and K_{dim} previously reported for wild type p62 UBA a K_d value of 47.1 ± 10 μM is produced. Using the same methodology, values of 10.3 ± 0.5, 7.1 ± 0.2 and 8.7 ± 0.4 μM were calculated for the S403A, S403D and S403E mutants respectively. An enhanced affinity for the three S403 mutants compared to wild type p62 UBA was observed. Again, there was a slight preference for the two phosphomimetic mutants over the S403A mutant.

7.2.3 Investigating how phosphorylation affects binding to ubiquitin using ESI-MS

ESI-MS titrations were performed as a complementary technique to back up the findings from the NMR titrations. Binding experiments using ITC were not possible as the data generated is unable to produce an accurate K_d value as it can not distinguish between the dissociation event and the binding event. ESI-MS titrations could be performed at concentrations low enough to minimize the effects of dimer dissociation. Moreover ESI-MS is able provide an insight into the stoichiometry of the interaction. All spectra were recorded in 25 mM ammonium acetate, pH 7.

7.2.3.1 ESI-MS titrations

ESI-MS titrations whereby ubiquitin was titrated into the various S403 p62 UBA mutants were performed. Up to 2 μM ubiquitin was titrated into 1 μM of each of the S403 mutants (as separate experiments). All three mutants produced nearly identical spectra with mass ions corresponding to the free S403 mutant, free ubiquitin and a 1:1 complex between the S043 mutants and ubiquitin (figure 7.20). No higher order oligomers or complexes with a different stoichiometry were observed. Mass ions corresponding to the complex showed weak intensities due to the interaction being largely hydrophobic. However, the intensities of the complex peaks for the two phosphomimetic mutants, S403D and S403E, are greater than those observed for the S403A mutant; indicating a stronger interaction for the phosphomimetic mutants. The complex peaks observed for the S403A mutant are barely above the noise level of the spectrum.

The peak corresponding to the +5 charge state for the dimer (m/z 2292, 2309 and 2315 for the S403A, S403D and S403E mutants respectively) was not visible at 1 μM in all cases confirming that the dimer is barely populated at this concentration and the K_d's produced reflect ubiquitin binding only.

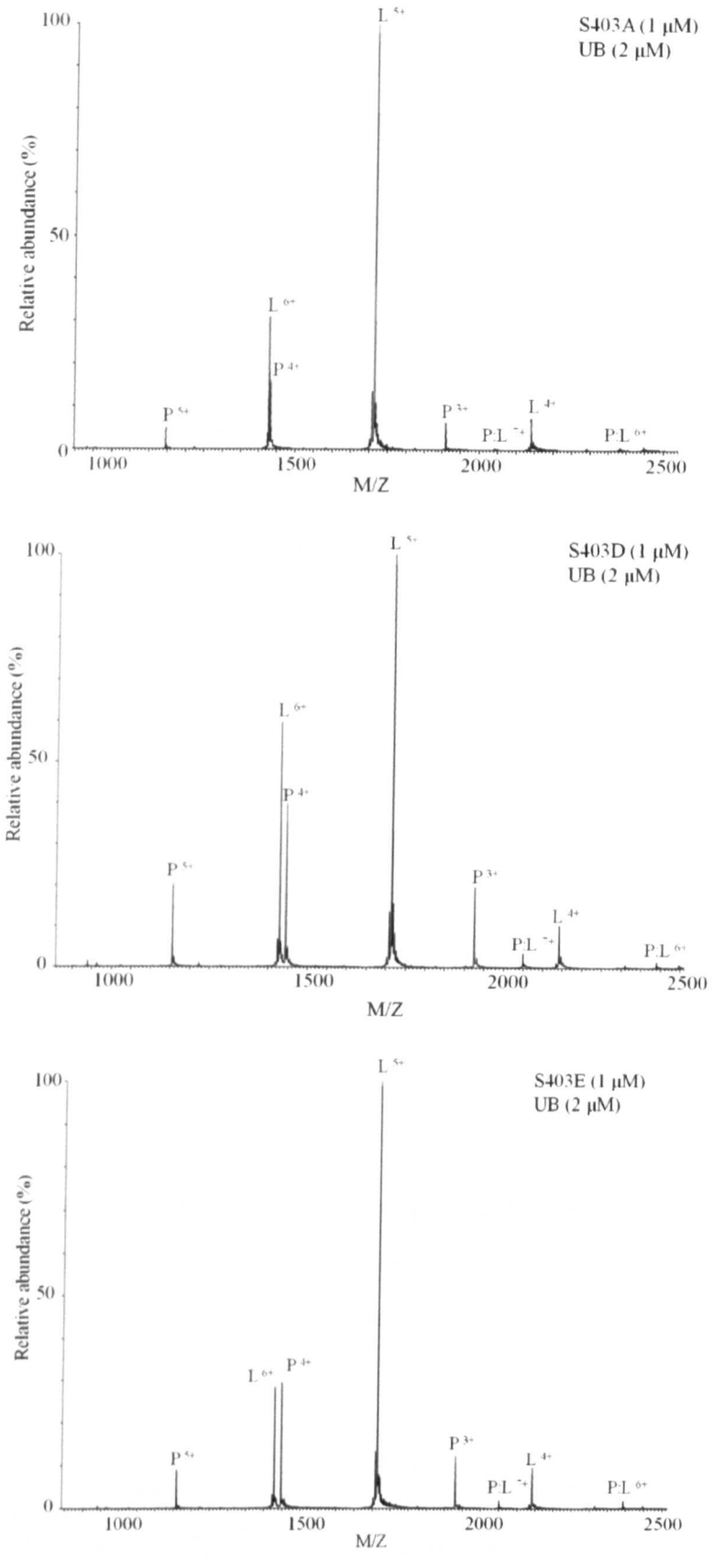


Figure 7.20 – The ESI-MS spectrum of 1 μM p62 UBA mutant A) S403A, B) S403D and C) S403E, mixed with 2 μM ubiquitin. The spectra were recorded in 25 mM ammonium acetate, pH 7. P represents the particular S043 mutant used in the experiment, L represents ubiquitin and P:L represents their complex.

By plotting [PL] against [L] a linear correlation is produced with a gradient of K_a is produced. From this the K_d values for each mutant can be calculated[222]. K_d values of 71.4 ± 5.5 μM, 45.5 ± 2.1 μM and 27.8 ± 0.8 μM were obtained for the S403A, S403D and S403E p62 UBA mutants respectively (table 6.3). These results, like the NMR reverse titrations, show a slight preference for the phosphomimetic mutants S403D and S403E over the S403A mutant. However, the K_d values are likely to be inaccurate due to the inability of the complex to remain intact once transferred to the gas phase. The complex is mainly stabilised by hydrophobic interactions between the two proteins, meaning that the mass spectrometer is inherently biased against complex formation in this case.

Table 7.3 – The K_d values for the S403 mutants and wild type p62 UBA as determined by ESI-MS titrations.

p62 UBA mutant	K_d (±) / μM
S403A	71.4 ± 5.5
S403D	45.5 ± 2.1
S403E	27.8 ±0.8

7.2.3.2 Competitive binding studies between the S403 mutants and using ESI-MS

The reverse titration studies by NMR confirmed that all three S403 mutants bind to the canonical Ile44/Val70 hydrophobic patch on ubiquitin. Since all three mutants bind to the same patch on ubiquitin, competitive binding studies where all three S403 p62 UBA mutants, which have different molecular weights, were used in a single experiment to determine which of the mutants has the greatest affinity. Despite the interaction between the S403 mutants and ubiquitin being hydrophobic, ESI-MS was the only technique capable of conducting this type of experiment as it can identify the small differences in molecular weight between the S403 p62 UBA mutants. By limiting the number of ubiquitin molecules compared to UBA binding sites, the S403 mutant with the greatest affinity could be established.

An experiment whereby all three S403 mutants and ubiquitin were mixed at equimolar concentrations (5 μM) was set up. An identical experiment using 1 μM of all the S403 mutants and 1 μM ubiquitin was also performed to minimize the effects of UBA dimerisation. Both experiments produced very similar spectrums, however the 5 μM spectrum showed more visible complex ions (figure 7.21).

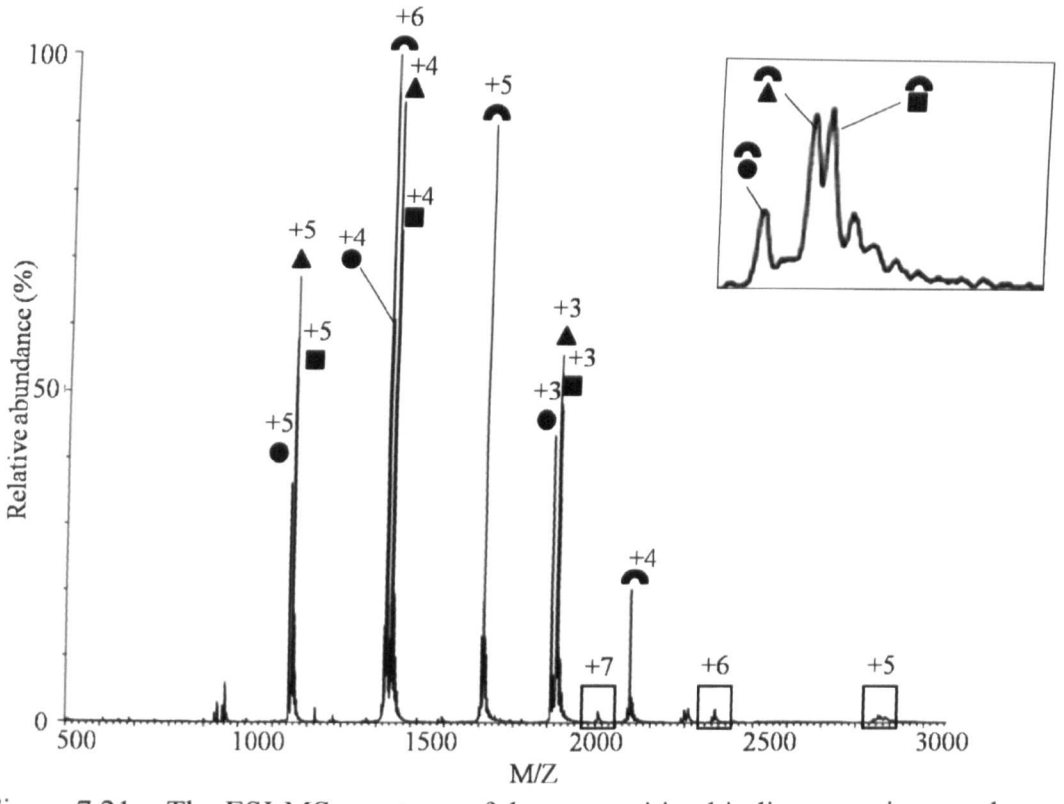


Figure 7.21 – The ESI-MS spectrum of the competitive binding experiment where 5 μM all S403 p62 UBA mutants was mixed with 5 μM ubiquitin. Spectra were recorded in 25 mM ammonium acetate, pH 7. ● is the S403A mutant, ▲ is the S403D mutant, ■ is the S403E mutant and ◠ is ubiquitin. ◓ , ◬ and ◙ are the S403A:Ub, S403D:Ub and S403E:Ub complexes respectively. The greater affinity of the phosphomimetic mutants S403D and S403E can be easily observed in this spectrum. Inset a zoom of the +6 charge state for all 3 complexes.

Similar to the individual binding experiments, complex formation was low given that the interactions were still relatively weak hydrophobic interactions being monitored in the gas phase. Peaks corresponding to the +3 to the +5 charge states of the free mutant UBAs were observed as well as the +5 to +6 for unbound ubiquitin. Peaks which correspond to the +5 to +7 charge state for all three complexes are also observed (inset figure 7.21), although the signal to noise ratio was low.

When the region of the spectrum where the complexes are formed (m/z 2000-3000) is enlarged there is an apparent preference for ubiquitin by the two phosphomimetic mutants S403D and S403E, compared to the dephosphorylated mutant S403A. The intensities corresponding to the S403D:Ub and S403E:Ub complex are almost double those observed for the S403A:Ub complex. From the intensities of the mass ions observed in the spectrum percentages of the free

mutant UBAs, and free ubiquitin can be calculated (table 6.3). From the percentages of ubiquitin and S403 mutant UBA bound within both the 5 μM and 1 μM reactions, a preference for the phosphomimetic mutants is clear. The percentages are only likely to be accurate if it is assumed that the maximum amount of complex has been formed. The amount of complex formed in the gas phase is unlikely to be accurate for these complexes because, as outlined earlier, the interaction is primarily hydrophobic.

Moreover, the S403 p62 UBA mutants were expected to ionise in the same way due to their similar size and solvent exposed surface area, although some slight variations in intensities of the free and complex ions for each mutant was observed. This would suggest that there are some small differences in the efficiency of ionisation, or that there are slight differences in the concentrations of the S403 p62 UBA mutants used in the experiment.

Table 7.4 – A table showing the percentage of the S403 mutants and ubiquitin which binds to form the complex in both the 5 μM and 1 μM eqimolar experiments

	5 μM eqimolar experiment		1 μM eqimolar experiment	
p62 UBA mutant	% UBA bound	% ubiquitin bound	% UBA bound	% ubiquitin bound
S403A	1.28	0.88	0.62	0.37
S403D	2.13	2.27	0.71	1.16
S403E	2.49	2.19	0.8	1.28

7.2.4 Investigating phosphorylation using GST pull downs and biological assays

All of the binding studies detailed in this chapter are to monoubiquitin rather than polyubiquitin chains. Full length p62 has been shown to have a greater affinity for polyubiquitin chains[104], thus it was important to determine some binding affinities to longer polyubiquitin chains. Part of the study conducted by Matsumoto *et al* used immunoprecipitation assays to detect the amount of polyubiquitin proteins that coimmunoprecipitated with the S403A and S403E mutants. The results of these experiments showed that the S403E mutant bound more ubiquitinated proteins compared to wild type; whereas, the S403A mutant bound slightly less than wild type.

To confirm the enhanced affinity of the S403E mutant, Miss Alice Goode conducted GST pull down assays on the S403A and S403E mutants. These mutations were investigated in both the isolated p62 UBA and full length p62 constructs. Binding was investigated using Lys63 linked polyubiquitin chains as well as monoubiquitin. The results of these pull down experiments showed a slight preference for ubiquitin by the S403E mutant, as indicated by a more intense band in the gel which represents the bound complex.

The effects of phosphorylation in specific PDB mutants, namely G425R, P392L, I424S and M404V, are also currently being investigated using GST pull down assays. Further biological assays are being conducted in cells on the PDB mutants by Miss Alice Goode. These assays aim to probe the effects of phosphorylation on the PDB mutants *in vivo*. The Layfield group has exclusive rights to the monoclonal antibody developed by Matsumoto *et al* that specifically recognizes S403-phospho p62. This antibody will determine if phosphorylation of the UBA occurs in the PDB mutants and how much of the phosphorylated form exists *in vivo*.

The effects of phosphorylation in the I424S PBD mutant are also being explored using biophysical techniques. Early signs show that the combined effect of the PDB mutation and phosphorylation markedly decreases the affinity of the interaction. For example, the I424S mutant alone was shown to have a K_d of 112.2 ± 5.1 μM, a value which is much weaker than the wild type p62 UBA (40 ± 10 μM). However, when the S403E mutation is also inserted to the I424A mutant a K_d of 198.4 ± 6.2 is calculated, highlighting an even weaker interaction for the double mutant (Miss Alice Goode, unpublished data). This is quite odd considering phosphorylation in the wild type p62 UBA shows an enhanced affinity. Moreover, the I424S mutant was shown to be a 5 fold weaker dimer with a K_{dim} of 19.8 ± 1. By inserting the S403E mutation into the I424S mutant the dimer becomes considerably weakened (K_{dim} 79.1 ± 3.3). The double mutant therefore shows a 20 fold decrease in K_{dim} relative to wild type (Miss Alice Goode, unpublished data). Again weaker dimers have been shown to have a higher affinity for ubiquitin. The I424S data highlights the interplay between different factors which combine to regulate the affinity. In the case of the I424S mutant thermal stability is likely to also have an effect.

7.3 Discussion

The function of phosphorylation as a regulatory PTM is well documented[246]. In the simplest of models a phosphorylated form and dephosphorylated form exist in equilibrium, whereby one form is biologically active. Complications occur when proteins are phosphorylated at several different sites, thereby producing various phosphoisoforms, each with a potentially different role *in vivo*. Similar to protein ubiquitination, multiphosphorylation of a single polypeptide chain can occur, generating a greater number of phosphoisoforms[258]. The study by Matsumoto *et al* identified eight different phosphorylation sites in the full length p62 protein; highlighting, that p62 has the potential for several biologically active phospho isoforms. Many of the phosphorylation sites on the p62 are found in regions linked to protein degradation pathways, therefore different phospho isoforms could regulate the p62s ability to shuttle proteins between different degradation pathways.

Since phosphorylation has the ability to alter the affinity of the interaction between the p62 UBA and ubiquitin, it is conceivable that phosphorylation at other sites in the full length protein could alter the affinity of p62 to bind to other proteins. Perhaps the most intriguing of the p62 phosphorylation sites is Ser332 which is located in the LIR and links p62 to autophagy. It has recently been shown that phosphorylation of the LIR in the optineurin protein, at residue Ser177, increases the affinity optineurin has for MAP-LC3 and subsequently increase the autophagic clearance of cytosolic Salmonella[259]. It is also plausible that other LIR containing proteins, for example NBR1, will have enhanced binding to MAP-LC3 when phosphorylated. NBR1 shares the same domain architecture as p62, has a role in selective autophagy and functions as a scaffold protein. Moreover, NBR1 contains a Ser as the first residue in the LIR, which corresponds to Ser332 in p62.

Some of the other phosphorylation sites in the p62 have also begun to be investigated. The cyclin dependent kinase cdk1 has recently been shown to phosphorylate p62 at both Thr269 and Ser272, both of which are located in the first PEST domain[260]. Disordered regions of proteins have been shown to have important roles in cell signaling pathways, with PEST sequences reportedly acting as internal degradation signals[137,138]. Phosphorylation in PEST domains have been

reported to influence ubiquitin mediated protein degradation[138]. The activity of cdk1 is regulated by the levels of cyclin B1 and phosphorylation of these sites on p62 has been shown to be crucial for control in the transition of cells through mitosis[260]. Interestingly, phosphorylation in the p62 PEST sequences was shown to have an effect on cyclin B1 stability. Cyclin B1instability is inactivated by phosphorylation of p62 at Thr269 and Ser272. Furthermore, cyclin B1 provides another link between the p62 and the UPS via Lys11 polyubiquitination. Cyclin B1 is ubiqutinated with Lys11 chains by the APC and is subsequently degraded by the proteasome. Although the exact molecular details of the interplay between the various degradation pathways remains unknown, p62 appears to have a central role in determining which proteins enter which degradation pathway, with the different pathways activated in response to specific cellular triggers.

Recently ubiquitin binding by the adaptor protein syntenin-1 was shown to be regulated by phosphorylation[261]. The two termini of syntenin-1 are required for the interaction with ubiquitin. Head-to-tail syntenin-1 dimers, mediated by two central PDZ domains, permit high affinity binding to two ubiquitin molecules. Synentin-1 dimerisation thus offers high avidity binding to ubiquitin. Dissociation of the syntenin-1:ubiquitin complex occurs when the ubiquitin binding motif in syntenin-1 is phosphorylated. The interaction with ubiquitin is therefore tightly regulated by both dimerisation and phosphorylation[261]. A similar control mechanism could also be employed by the p62 UBA, although dimerisation would downregulate and phosphorylation would upregulate ubiquitin binding.

Although there are no other reported examples of phosphorylated UBAs to date, the p62 UBA is most likely to be the first of many. Most ubiquitin:UBA interactions have been shown to be very weak with affinities that are not physiologically relevant. It is possible that phosphorylation could increase the affinity of other UBA interactions to high affinity interactions which promote the activation of specific cell signaling pathways. Unlike the p62 UBA, not all UBAs are found in proteins which can form higher order oligomers and are therefore unable to increase the affinity of the interaction via avidity effects. It therefore seems likely that other factors such as phosphorylation or linkage specificity play a role in regulating the interaction between other UBAs and ubiquitin in order to activate downstream cell signalling.

7.4 Conclusions

The wild type p62 UBA had been shown to form a highly stable dimer with a K_{dim} of 4.1 μM[141]. Studies in chapter 6 of this thesis revealed that dimerisation was able to regulate the affinity for ubiquitin. A recent study also highlighted the phosphomimetic mutant S403E as having a greater affinity for ubiquitin than wild type full length p62. The main aim of the experiments conducted in this chapter was to investigate the role of phosphorylation had in regulating the affinity for ubiquitin. The phosphomimetic mutants S403D and S403E were generated as well as a control S403A mutant to represent the dephosphorylated form.

In order to study any changes in affinity associated with the S403 mutants, it was first important to establish that the monomer-dimer equilibrium has not been altered by the introduced mutation. Dilution studies using ESI-MS, NMR and ITC similar to those performed for the helix 2 dimerisation interface were also conducted for the S403 mutants. The data from the dilution studies using different techniques correlated well to each other. Evidence for two species in equilibrium was detected in all cases. The NMR data highlighted the relative populations of monomer and dimer to be very similar to those observed for wild type protein at low concentration (10 μM). Another crucial observation determined from the NMR data was that all the mutants showed a spectrum with a similar dispersion of chemical shifts to wild type protein. The mutants were likely to possess a native-like protein fold. When examined closer only a few residues, all of which were in close proximity to the mutation site, showed chemical shift deviations compared to the wild type p62 UBA. The idea of a native-like fold was confirmed by the ESI-MS dilutions, which highlighted the solvent exposed surface area to be similar to wild type. Moreover, values for K_{dim} produced by ITC dilution studies were in excellent agreement with the K_{dim} value of 4.1 ± 0.6 μM observed for wild type p62 UBA. Taken together the data indicates that the three S403 mutants form stable dimers with similar K_{dim} values to each other and to wild type p62 UBA.

With the confirmation that mutations at position 403 having little effect on the structural integrity of the dimer attentions were turned to binding studies using NMR and ESI-MS. A forward NMR titration whereby unlabelled ubiquitin was titrated into ^{15}N-S403D p62 UBA showed saturation occurred after the addition of 10 molar equivalents of ubiquitin. Wild type experiments showed saturation after

20 molar equivalents, meaning the S403D mutant an increased affinity for ubiquitin compared to wild type p62 UBA. The reverse NMR titrations, whereby unlabelled mutant UBA was titrated into ^{15}N-ubiquitin, highlighted that ubiquitin had a preference for the phosphomimetic mutants, S403D and S403E p62 UBA. Saturation occurred for these mutants at 1 molar equivalent; whereas, saturation for the S403A and wild type p62 UBA occurred at 2 and 4 molar equivalents respectively. Larger chemical shifts were observed in the binding patch on ubiquitin when bound to the phosphomimetic mutants, again providing evidence for a stronger interaction. However, when K_d values for the interaction were calculated the data revealed only small differences in affinity between the mutants.

ESI-MS titrations where ubiquitin was titrated into the S403 p62 UBA mutants visually showed a clear preference for the phosphomimetic mutants over the S403A mutant. This finding was backed up by the competitive binding study, whereby the mutant UBAs were mixed together. The competitive ESI-MS showed a clear preference for the phosphomimetic mutants over the S403A mutant. Inaccuracies in the K_d values were likely to be caused by the gas phase and that the is principally hydrophobic.

Table 7.5 – A summary of the calculated K_{dim} and K_d values for wild type p62 UBA and the S403 mutants using various biophysical techniques. An ESI-MS experiment was not completed for wild type p62 UBA, therefore a K_d could not be calculated.

p62 UBA mutant	K_{dim} (±)/ μM (ITC)	K_d (±)/ μM (fwd NMR)	K_d (±)/ μM (rev NMR)	K_d (±)/ μM (ESI-MS)
Wild type	4.1 (±0.6)	40 (±10)	47.1 (±10)	
S403A	3.5 (±0.5)	N/A	10.3 (±0.5)	71.4 (±5.5)
S403D	6.7 (±0.6)	39.1 (±4.7)	7.1 (±0.2)	45.5 (±2.1)
S403E	4.2 (±0.5)	N/A	8.7 (±0.4)	27.8 (±0.8)

Overall small differences in affinity were observed for the S403 mutants only. This is similar to the effects to the PDB mutations, in which subtle decreases, rather than increases, in affinity occur. The small differences in affinity can be amplified by oligomeric p62 *in vivo* because of the avidity effect (discussed in chapter 9). The potential for arrays of Ser403 phosphorylated UBAs to be presented to long polyubiquitin chains is likely when oliogomerisation occurs via the N-terminal PB1 domain. The avidity effects enhanced further due to the dimerisation of the p62 UBA itself. The binding studies in this chapter were only conducted using monoubiquitin which precludes the analysis of the role of avidity. In addition to this the role of linkage specificity exhibited by the phosphorylated p62 UBA has not been investigated by the experiments completed in this chapter.

The phosphorylation site found in the UBA, Ser403, is found adjacent to the MGF motif in loop 1 which binds directly to the hydrophobic Ile44/Val70 patch on ubiquitin. Therefore the phosphorylated form of the UBA could induce small structural changes in the binding motif to produce a better fit between the molecules. Structural perturbations were observed by NMR for residues surrounding the mutation site but a significant structural rearrangement of the whole UBA was not seen. This is consistent with local changes to the binding site to produce a surface that has a greater complementarity to ubiquitin. By plotting CSPs to the surface of the ubiquitin, a more extensive binding patch and tighter binding interface are observed for the two phosphomimetic mutants, despite affinity differences of only 2-3 fold.

Mutating to Ala, as in the S403A mutant, substitutes a polar amino acid for a hydrophobic one. Since Ser403 is located in the hydrophobic patch which binds to ubiquitin, the presence of an additional hydrophobic residue could facilitate a slightly stronger interaction between the S403A UBA and ubiquitin. This possibly explains why the S403A mutant does not act as a perfect non-phosphorylated control. Mutating to the S403D and S403E mutants alters the charge, which also has the potential to favour electrostatic interactions at the binding interface between the UBA and ubiquitin and could be a reason why phosphorylation has been shown to enhance the interaction.

Taken together, experiments in this chapter, as well as those conducted by Matsumoto *et al*, phosphomimetic mutants of the p62 UBA have shown that the phosphorylated form of the UBA has an enhanced affinity for monoubiquitin both *in vitro* and *in vivo*. *In vivo* phosphorylated p62 UBA has been shown to target more polyubiquitinated proteins to sequestosomes for their degradation by autophagy. Although the extent of phosphorylation *in vivo* remains unknown, it seems likely that the proposed equilibrium between the phosphorylated and the dephosporylated forms of the p62 UBA exists and that the degree of phosphorylation occurs in response to the polyubiquitinated protein level. Phosphorylation at Ser403 therefore seems to be an important regulator in determining the fate of ubiquitinated proteins.

7.5 Future work

The data collected in this chapter clearly indicate some structural alterations in the MGF loop for the phosphomimetic S403D and S403E p62 UBA mutants. These alterations were not observed for the S403A mutant. An extension to the project could be to investigate the changes caused by mutation in the phosphomimetic mutants further. A complete set of NMR experiments, including 3D triple resonance experiments, could be used to generate structures of the phoshomimetic mutants in both the free and bound to ubiquitin forms. Molecular modelling experiments using the phosphomimetic mutants and ubiquitin could be conducted to back up the NMR structure. Both of these techniques would provide the exact details of how the binding interface has been affected by mutation.

Another extension to the work outlined in this chapter could be to perform both NMR and ESI-MS titrations using the phosphomimetic mutants and longer polyubiquitin chains. GST pull downs had highlighted the p62 UBA an increased affinity for polyubiquitin chains over monoubiquitin[170,171], therefore larger changes in affinity might be observed by longer chains, especially as only small changes were observed in monoubiquitin binding. Longer chains would also be expected to demonstrate higher affinity interactions in the presence of oligomeric p62 *in vivo,* as a result of the combined synergistic strength of the individual interactions, known as avidity. Such avidity effects would be amplified if the individual interactions were with the phosphorylated form of the p62 UBA. Additionally chains of different linkage could be used to probe linkage specificity. A dramatic increase in the affinity for Lys48 linked polyubiquitin chains, as well as a slight increase for Lys63 linked polyubiquitin chains, was observed by Matsumoto *et al* in GST pull down assays using the S403E mutant p62.

A final extension to the project could be to investigate the other phosphorylation sites. The phosphorylation site found in the LIR at position Ser332 is perhaps the most obvious first choice, especially as phosphorylation has been observed in the optineurin LIR and the affinity for MAP-LC3 has been shown to increase in its phosphorylated form[259].

8.0 The p62 protein mediates the formation of a ternary complex between ubiquitin and MAP-LC3

8.1 Introduction

8.1.1 Autophagy

Autophagy is a degradation pathway which is highly conserved among eukaryotes (figure 8.1). The process involves sequestering misfolded/aggregated proteins and damaged organelles to autophagosomal vesicles for their degradation by lysosomes[262,263].The down regulation of autophagy has been linked to neurodegeneration and cancer as well as regulating immunity[264,114,265]. Broadly speaking autophagy can be divided into three main types microautophagy, chaperone-mediated autophagy (CMA) and macroautophagy depending on the delivery route of the material to a membrane bound organelle known as an autolysosome[263]. Macroautophagy is considered to be the most common route, as a well as being the most extensively studied and is referred to as autophagy hereafter.

Autophagy is regulated by a protein kinase known as Tor[266]. Tor is responsible for the phosphorylation of proteins which inhibit the initiation of autophagy. When these proteins are dephosphorylated autophagy is activated[266]. In autophagy, a phagophore is formed by the nucleation of an isolated membrane. The exact origin of the membrane involved in phagophore formation is not yet clearly known, however examples have been shown for its origin in various membranous structures such as the endoplasmic reticulum and golgi bodies[267]. When the phagophore is sealed it becomes a double membrane and is termed an autophagosome. The autophagosome then undergoes a maturation process culminating in fusion with a lysosome. This structure, now termed an autolysosome, is capable of degrading the intracellular material. The breakdown occurs via acidic hydrolases found within the lysosome[263,268]. Autophagy is an essential process during times of cell starvation enabling the cell to use the degraded materials as sources of energy[263].

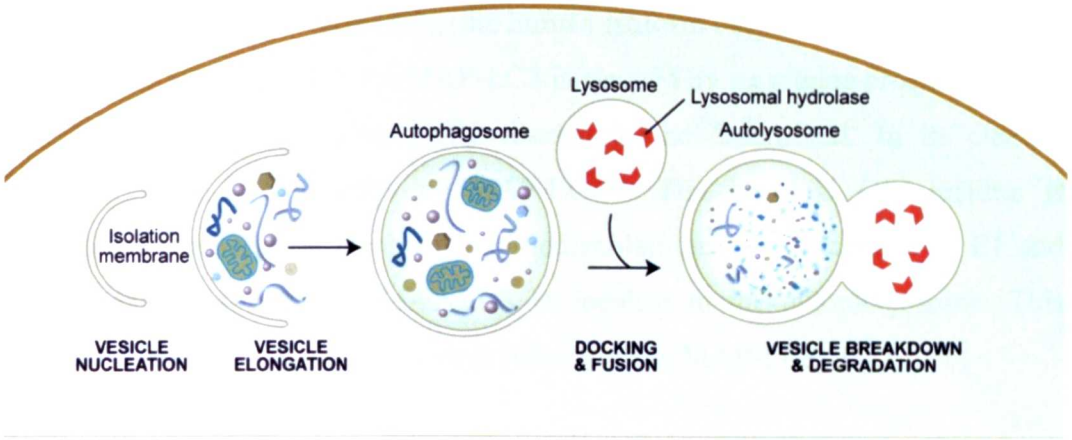


Figure 8.1 – Schematic representation of autophagy in the cell. The formation of the autophosome, its maturation and the degradation of its contents are depicted.

8.1.2 Microtubule-associated protein 1 light chain 3 (MAP-LC3) protein

MAP-LC3$_B$ (the human isoform used throughout this study) is a human homologue of the yeast protein Atg8 which is essential for autophagy. MAP-LC3 proteins belong to the microtubule-associated protein (MAP) family and are small ubiquitin-like proteins with molecular weights of around 15 kDa. Within the autophagic pathway, MAP-LC3 has been shown to have a crucial role in the formation of autophagosomes with the N-terminal α helices being fundamental in membrane fusion[269]. There are 6 human homologues of the yeast Atg8, MAP-LC3$_A$, MAP-LC3$_B$, MAP-LC3$_C$, GABA$_A$ receptor associated protein (GABARAP), GABARAPL1 and GABARAPL2 (which is also known as GATE-16). Little is known about GABARAPL1 and GABARAPL2; however, the 3 isoforms of MAP-LC3 and GABARAP have been studied more extensively[262]. Structurally, human MAP-LC3 proteins are similar to other mammalian homologues despite having low sequence homology. MAP-LC3 proteins possess two subdomains and are comprised of 4 α-helices and a central mixed β-sheet (figure 8.2). The β sheet contains 4 strands denoted S1 to S4. S1 and S4 are parallel strands located in the centre of the sheet, whereas S2 and S3are antiparallel to S1 and S4 respectively. Helices 1 and 2 form the N-terminal subdomain. Helices 3 and 4 and the β sheet form a ubiquitin-like fold and is referred to as the C–terminal subdomain. The two subdomains subsequently fold into a single compact globular structure driven by hydrophobic patches within each subdomain[270].

In its full length form MAP-LC3$_B$ (the human isoform used throughout this study) is known as ProMAP-LC3. ProMAP-LC3 is cleaved by a cysteine protease known as Atg4B leaving an exposed Gly residue at the C-terminal. In its cleaved cytosolic form it is known as MAP-LC3-I. The terminal Gly residue is subsequently conjugated to phosphatidylethanolamine via the actions of E1 and E2 like enzymes, enabling MAP-LC3 to localise to the autophagosome. This membrane bound lipidated structure is referred to as MAP-LC3-II[271].

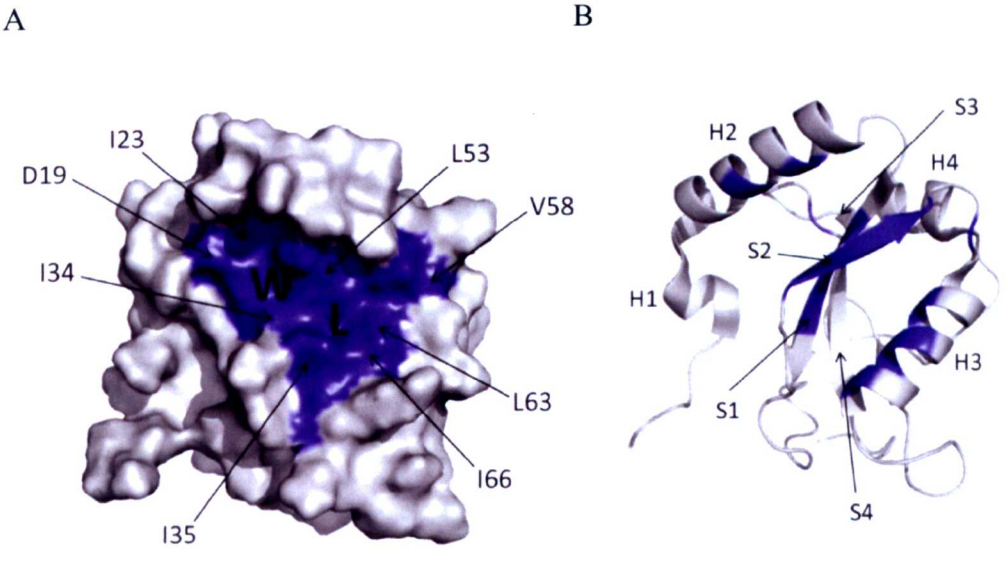


Figure 8.2 – The structure of the MAP-LC3$_B$ protein (PDB ID 1V49) A) A surface plot of MAP-LC3$_B$ depicting the two hydrophobic pockets (coloured blue) which bind to the conserved WXXL motif in the LIR of the p62. The pockets are labelled according to the residue which they bind. B) Ribbon representation of the MAP-LC3$_B$ protein. The hydrophobic residues are again highlighted in blue.

8.1.3 The role of p62 in autophagy

The full length p62 protein acts as a molecular scaffold which links target proteins to two degradative pathways: autophagy and the ubiquitin proteasome system (UPS). The UPS is involved in the disposal of short lived proteins; whereas long lived proteins and organelles which have become damaged are removed by autophagy. Ubiquitin mediated selective autophagy involves the formation of an autophagosome which contains ubiquitinated proteins bound to p62. Ubiquitinated proteins are often delivered to the proteasome for degradation but are also commonly found in autophagosomes the p62 where they are degraded by autophagy. The p62 is able to transfer target ubiquitinated proteins to different routes of degradation. There is mounting evidence for cross talk between autophagy and the UPS[272,273], although the molecular mechanisms responsible for this remain unknown. Also rather interestingly the p62 protein is degraded by autophagy[274] but is protected from degradation by the proteasome by its C-terminal UBA[143].

The p62 protein contains a 22 residue MAP-LC3 interacting region (LIR) which binds directly to the MAP-LC3 protein, linking the p62 protein to the autophagy pathway[116, 117]. The p62 LIR is located between residues 332 and 345, with a conserved WXXL motif used for direct interaction with the p62[116, 275]. This motif, WTHL as it is found in the p62 protein, spans residues 338-341 and has been shown to exhibit an extended β conformation. An intermolecular parallel β-sheet is formed between the WTHL motif of the p62 and the β2 of the MAP-LC3 C-terminal subdomain. The Trp338 side chain interacts with the hydrophobic pocket formed at the interface between the two subdomains of the MAP-LC3; whereas, the Leu341 sidechain binds to the hydrophobic pocket on the C-terminal subdomain[275]. The p62 LIR also contains an acidic patch consisting of three consecutive Asp residues at positions 335 to 337. This patch has been shown to interact with basic residues at the N-terminal of the MAP-LC3[117].

The p62 protein is very similar in domain structure to neighbour of BRCA gene 1 (NBR1) protein, with each containing an N-terminal PB1 domain, a C-terminal UBA domain and an LIR (figure 8.3, A). The LIR and UBA enable these two proteins to bind to both MAP-LC3 and ubiquitin. Both proteins have been shown to exhibit roles as adaptor proteins in selective autophagy. More recently a role for

ubiquitin has been discovered in selective autophagy, since these adapter proteins are involved in the delivery of polyubiquitinated misfolded substrates which are degraded by autophagy[116,276]. Both are capable of polymerising or aggregating via their PB1 domains and can recognise substrates required for selective autophagy[277]. It has been recently proposed that self oligomerisation via the PB1 domain is a prerequisite for targeting to the autophagosome, rather than MAP-LC3 binding. The proteins are subsequently incorporated into the autophagosome in a process that is most likely MAP-LC3 dependent although the exact details have yet to be elucidated[278]. Despite their similarities, some diversity between p62 and NBR1 has been observed, p62 has been shown to possess a specific role in cell protection by targeting invading bacteria to the autophagy pathway, unlike NBR1[134].

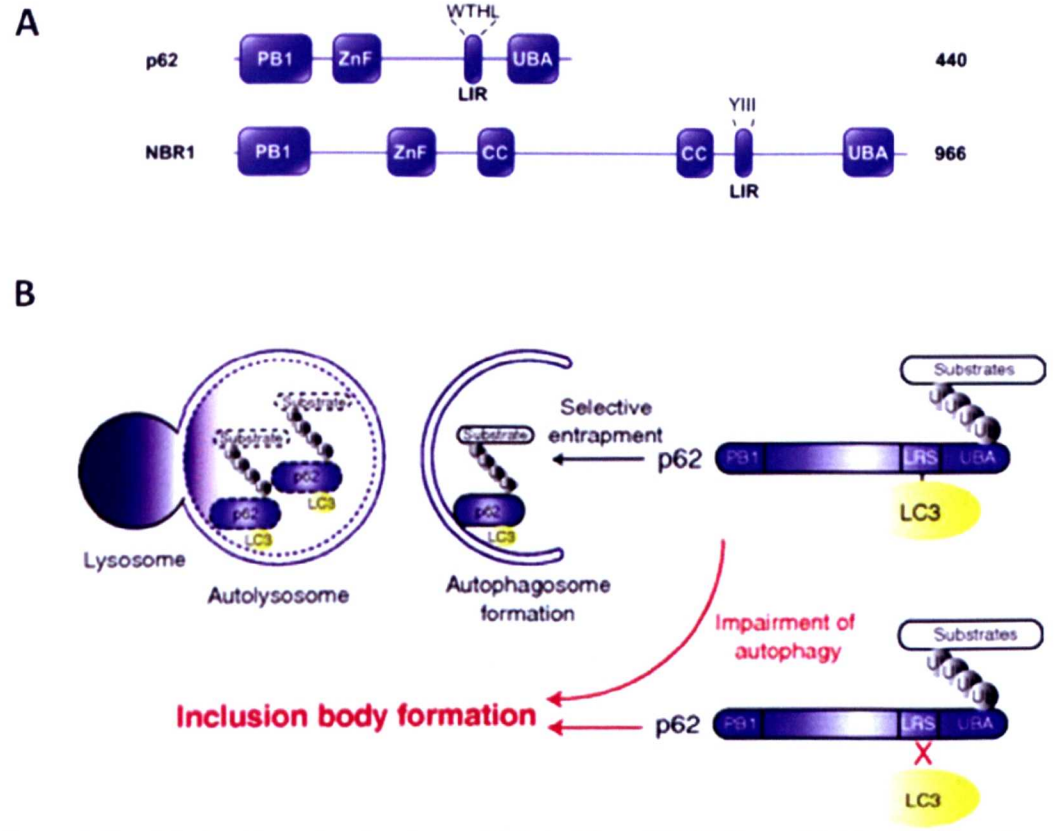


Figure 8.3 – A) the domain structure of the p62 and NBR1 scaffold proteins. The NBR1 protein possesses coiled coil motifs in addition to the PB1, ZnF, UBA and LIR which are shared by both proteins. Both of NBR1 and p62 are capable of binding to both ubiquitin and MAP-LC3 and have roles in selective autophagy. The crucial interacting residues of both LIRs are provided. Taken from Kirkin *et al*[279]. B) a schematic model for selective autophagy mediated by p62. Taken from Ichimura *et al*[117]. The p62 binds to target proteins and both proteins become trapped within the autophagosome and are subsequently degraded.

As more LIR regions in different proteins have been discovered some variability has been shown. The WXXL motif has now been modified to a more general θXXΓ sequence, where θ represents an aromatic amino acid and Γ represents a hydrophobic residue (table 8.1). It has also been noticed that negatively charged residues and amino acids which are capable of being phosphorylated are also common features to LIRs. Since the residues are only partially conserved within different LIRs, investigations into how critical the amino acids are within the θXXΓ sequence[280].

LIR containing protein	LIR amino acid sequence	LIR location
P62	SGGDDDWTHLSS	332-345
NBR1	SASSEDYIILPE	727-738
Nix W36	AGLNSSWVELPM	30-41
Nix W140	SADWVSDWSSRPENIP	137-152
Atg4B	TLTYDTLRF	5-13
Sc Atg19	ALTWEEL	409-415

Table 8.1 – A comparison of the LIR sequences in different proteins. The aromatic and hydrophobic amino acids are shown in red and blue respectively.

8.1.4 Summary and aims

The separate interactions between the p62 LIR and MAP-LC3 and the p62 UBA and ubiquitin have both been reported by NMR[275,140]. However the interaction with MAP-LC3 was provided by using a p62 peptide corresponding to the LIR region only. Since the full length p62 protein functions as a scaffold protein it is highly likely that the p62 can bind to multiple binding partners simultaneously to regulate different signalling pathways. Moreover, the p62 protein provides a molecular link between ubiquitination and autophagy so it therefore seemed logical to see if a p62 mediated ternary complex could be formed between p62, MAP-LC3 and ubiquitin (figure 8.4). We also aim to investigate if there is an allosteric relationship between the two binding proteins.

To date no structural work has been conducted on the full length p62 protein or any of the potential multimeric protein complexes the p62 protein could form. In order to gain an insight into the p62s ability to regulate cross talk between the UPS and autophagy, binding studies using NMR spectroscopy were completed. We aim to rationalise the structural basis for recognition of both MAP-LC3 and

ubiquitin by the p62. Insights into the molecular recognition of one protein whilst the other is pre-bound, is crucial to understanding the p62s ability to regulate of the two degradation pathways. It was also unknown if any contacts between MAP-LC3 and ubiquitin occur once simultaneously bound to the p62. NMR would be able to identify if such contacts are formed. In order to back up any findings by NMR, binding experiments using ESI-MS were also conducted. ESI-MS has the advantage of using low sample concentrations and quick data collection times. Moreover, it would be possible to detect if an interaction between the MAP-LC3 protein and ubiquitin occurred, for which NMR is a unique tool.

The full length p62 had previously been shown to be difficult to purify, due to a large unstructured region between the ZZ domain and the UBA which encodes two PEST domains. PEST domains are thought to function as internal degradation signals[137,138], although this is unconfirmed in the p62. Therefore a fragment of the p62 extending back from the C-terminal which encompassed both the UBA and the LIR, but minimised the unstructured region was engineered. This construct, denoted p62 300-440, consisted of eighty seven amino acids of unstructured protein and the folded UBA domain. The p62 300-440 construct was mutated to a C331S variant to eliminate the potential to form an intermolecular disulphide bond.

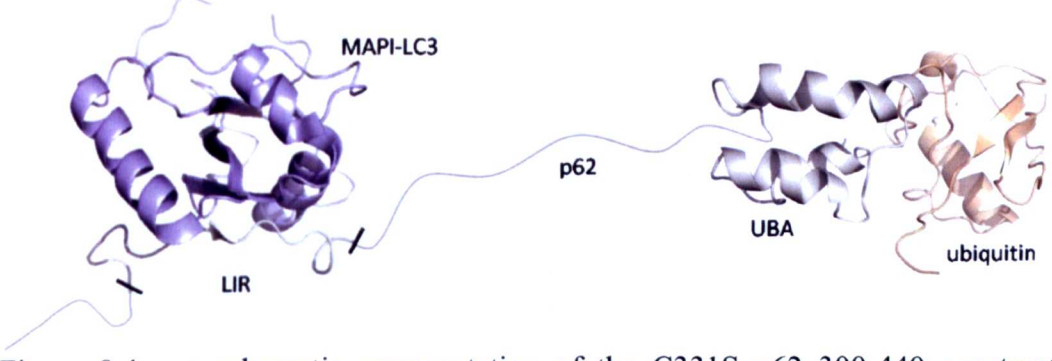


Figure 8.4 – a schematic representation of the C331S p62 300-440 construct bound to the MAP-LC3 and ubiquitin. The image was made from the MAP-LC3 protein complex to the p62 LIR (PDB ID 2K6Q) and the ubiquitin-p62 UBA complex. The ubiquitin-p62 UBA complex is a representation made from the PBD's of ubiquitin and the bound monomer (PDB ID's 1UBQ and 2JY8). Black lines are used to define the LIR in the schematic.

8.2 Results

8.2.1 Purification of the C331S p62 300-330 reveals partial degradation

The C331S p62 300-440 construct was able to be purified to a low yield (approximately 3 mg / Litre). A fully homogenous sample was not able to be obtained as the sample progressively degraded throughout the purification, despite the addition of a protease cocktail at each stage of purification. The degradation was able to be visualised by SDS PAGE; however, the intensity of the fragmented bands was weak compared to the intensity of the full length protein (figure 8.5). It is possible that degradation is an intrinsic property of the sequence, since the C331S p62 300-440 fragment encodes one of the two PEST domains.

Unfortunately there was no way to quantify the amount of degradation caused. The nanodrop which was used to check concentrations would detect the flourophores in the fragments as well as the full length C331S p62 300-440. It was difficult to determine exactly where the degradation was occurring as several recognition sequences for proteases were contained in the linker region. The gel suggests a cleavage close to the UBA as a band appears close to the 6.5 kDa marker. Since the intensities of the fragmented bands were weak it was decided that the level of degradation was not enough to alter protein concentration significantly. With this in mind it was decided to proceed with NMR experiments.

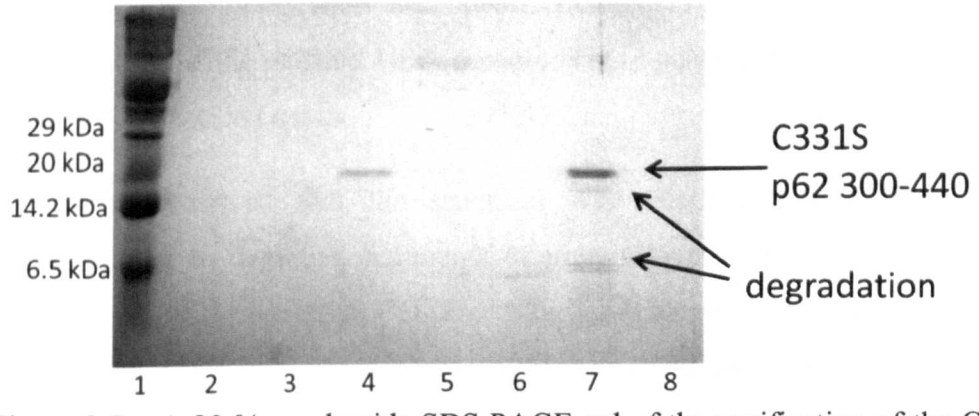


Figure 8.5 – A 20 % acrylamide SDS PAGE gel of the purification of the C331S p62 300-440 construct which had been in expressed in C41 (DE3) *E.coli.*. A low molecular weight marker was used as a size standard (lane 1). Gel filtration fractions 13, 14, 15, 16 and 17 are shown in lanes 2 -6 respectively. Desalt fractions 2 and 3 are shown in lanes 7 and 8. C331S p62 300-440 containing fractions are gel filtration fraction 15 and desalt fraction 2. The degradation has become more evident after desalting as shown by the appearance of faint bands beneath the main protein band.

8.2.2 NMR studies on the p62 mediated ternary complex between C331S p62 300-440, MAP-LC3 and ubiquitin

All NMR samples were prepared by dissolving lyophilised protein in 25 mM potassium phosphate, 25 mM NaCl, pH7 buffer. All spectra were recorded at 298K. Initial experiments to confirm the suitability of the ^{15}N-C331S p62 300-440 construct for NMR experiments were conducted prior to ^{1}H-^{15}N TROSY titrations, ^{1}H-^{15}N heteronuclear NOE experiments and 3D triple resonance experiments for protein backbone assignment.

8.2.2.1 The ^{1}H ^{15}N TROSY experiment confirms the C331S p62 300-440 contains both folded and unstructured regions

A 250 μM sample of ^{15}N-C331S p62 300-440 was prepared and it's conformation checked using a ^{1}H-^{15}N TROSY (figure 8.6). The ^{1}H-^{15}N TROSY spectrum shows two distinct regions of the C331S p62 300-440 protein, an unstructured region and the folded UBA. The unstructured region is represented by the boxed region in the centre of the spectrum (^{1}H dimension 7.9 – 8.5 ppm, ^{15}N dimension 112 – 128 ppm). These peaks show poor dispersion with a significant degree of overlap and are very strong in intensity. All of these features are characteristic of unstructured regions of proteins. In this construct, residues 300-387 correspond to part of the intrinsically disordered region of the p62. The well dispersed peaks on the outer most parts of the spectrum (generally outside the boxed region) correspond to the p62 UBA. These peaks show near identical chemical shifts to those reported for the free form of the isolated UBA domain. These peaks are significantly weaker than the unstructured peaks.

It was also noticed that the spectrum was relatively clean. Some peaks corresponding to degraded fragments of proteins were observed in the central lower region of the spectrum (^{1}H dimension 7.5-8.5 ppm, ^{15}N dimension 125-130 ppm) when the levels were increased. However, in agreement with the SDS PAGE analysis it was deduced that significant proteolysis of the ^{15}N-C331S p62 300-440 construct had not occurred.

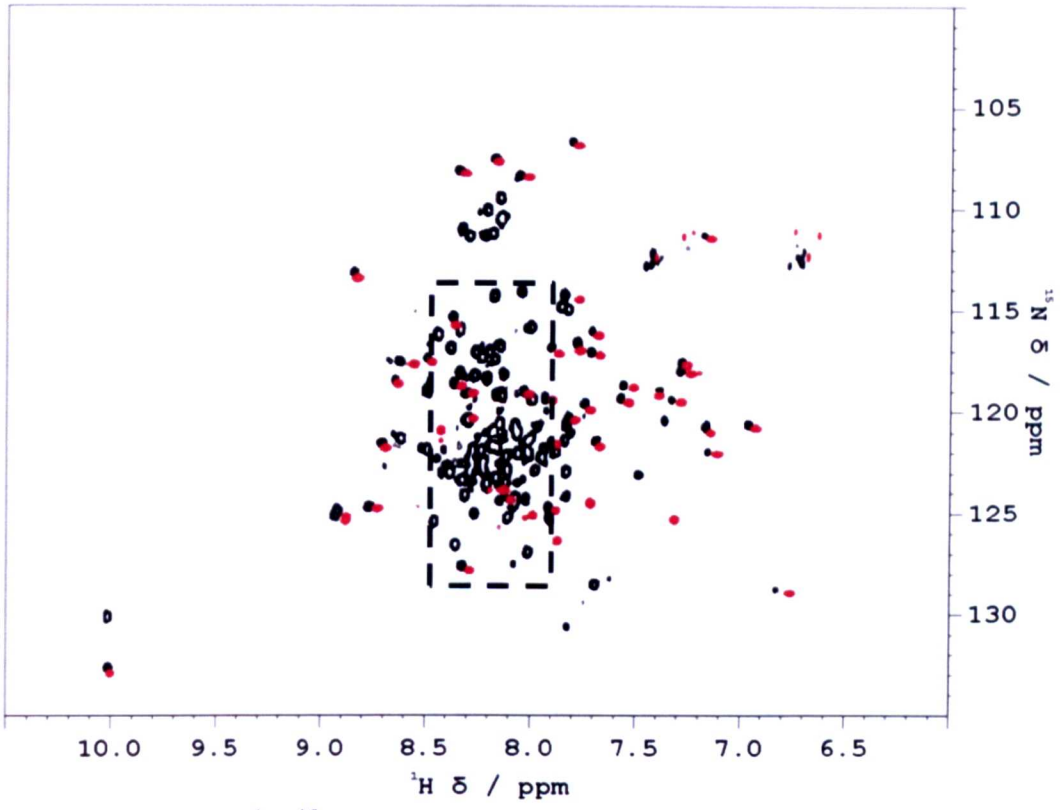


Figure 8.6 – The ^{1}H-^{15}N TROSY spectrum for the free C331S p62 300-440 construct (black) and the p62 UBA (red). The UBA peaks in both constructs overlay well. The central region containing the unstructured amino acids in the C331S p62 300-440 construct is shown by the dotted box. Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K. The well dispersed peaks on the edges of the spectrum belong to the folded UBA; whereas, the central highly overlapped region corresponds to the unstructured region located N-terminally to the UBA domain.

8.2.2.2 NMR titrations reveal binding of MAP-LC3 and ubiquitin is not allosteric

A titration involving the binding of unlabelled MAP-LC3 protein to the ^{15}N-C331S p62 300-440 construct then adding unlabelled ubiquitin to the MAP-LC3 bound form was conducted. A reverse titration involving adding ubiquitin first, then MAP-LC3 was also conducted. An initial attempt at a titration revealed that the lifespan of the ^{15}N-C311S p62 300-440 is only approximately 48 hours. After this time period the ^{15}N-C331S p62 300-440 shows substantial degradation. Attempts were made to shorten the length of the experiment by using the SOFAST pulse sequence on the ^{1}H-^{15}N TROSY; however, extensive T1 noise caused a large streak through the centre of the spectrum. This noise confirmed that this type of experiment was not beneficial to this particular protein. With this in mind the subsequent titrations were completed using the lowest number of

scans whilst maintaining signal to noise. This time reduction permitted the titrations to be started and finished on the same day in order to reduce sample degradation as much as possible. The titrations show that both proteins bind to the p62 and bind independently of occupation of the other binding site. The data also revealed that there is not an allosteric relationship between MAP-LC3 and ubiquitin. Both proteins bind to the ^{15}N-C331S p62 300-440 construct with similar affinities to those reported for the separate interactions, as saturation is reached at similar points in the titrations.

The addition of up to 250 μM unlabelled MAP-LC3 to 250 μM ^{15}N-C331S p62 300-440 resulted in concentration dependent changes in chemical shift. A mixture of fast and slow exchange was observed for the resonances which were perturbed upon MAP-LC3 binding. Slow exchange means that two peaks per residue are observed during the titration, one for the free form and one for the bound form. On the other hand, fast exchange visualises a population weighted average between the free and bound forms enabling the movement of peaks to be tracked. The LIR is just twelve amino acids in length, with the conserved binding motif consisting of just four residues. This means that only a small part of the ^{15}N-C331S p62 300-440 fragment is altered upon binding to MAP-LC3. Significant changes in chemical shift are observed for approximately fifteen residues, with no changes in chemical shift observed for residues previously assigned to the UBA (figure 8.7).

The interaction between the LIR of the p62 and MAP-LC3 is of a higher affinity than the interaction observed for the UBA and ubiquitin, with saturation occurring around 1:1 molar equivalents. This is consistent with the previous study between the p62 LIR peptide and MAP-LC3. It would therefore appear that using a longer fragment of the p62 does not dramatically alter the affinity the p62 has for MAP-LC3. Moreover, it agrees well with the interaction between the LIR1 of NBR1 and MAP-LC3 (K_d 0.62 ± 0.06 μM, determined using ESI-MS)[281]. The NBR1 protein is structurally and functionally very similar to p62; therefore, the mechanism of binding and the affinity for MAP-LC3 is likely to be similar.

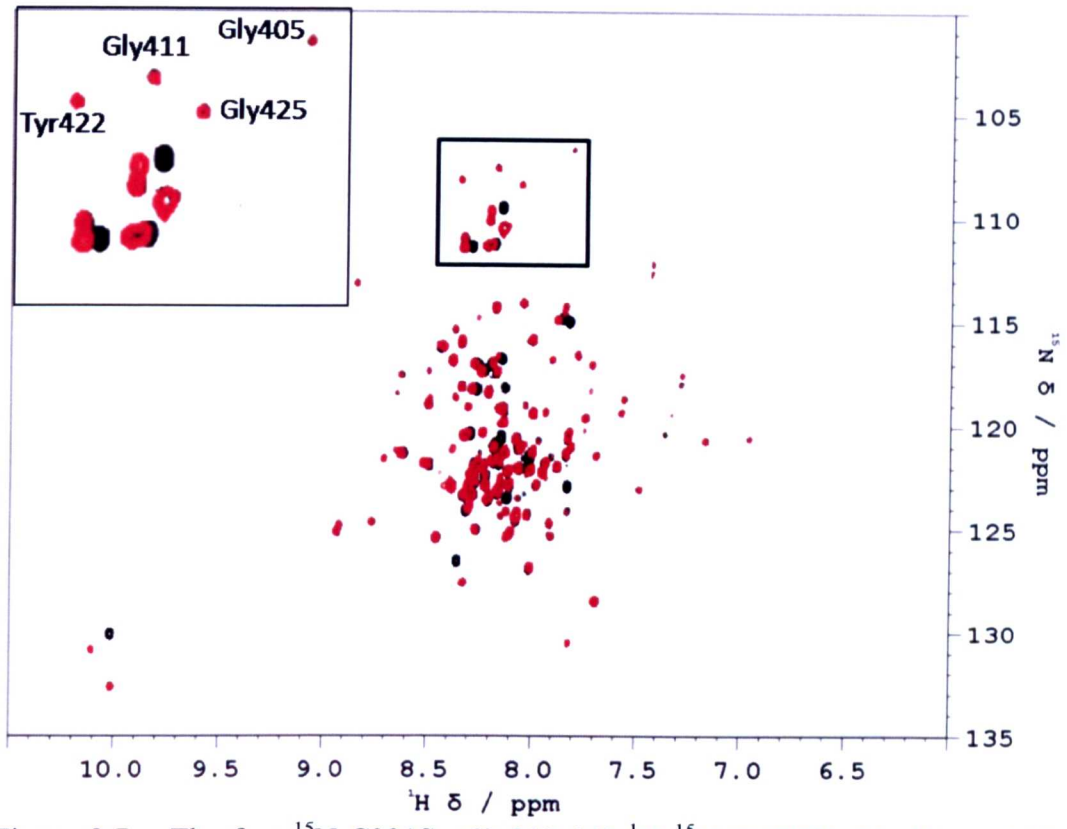


Figure 8.7 – The free ^{15}N-C331S p62 300-440 ^{1}H ^{15}N-TROSY (black) overlaid with the ^{1}H-^{15}N TROSY of the ^{15}N-C331S p62 300-440 bound to MAP-LC3 (red). Inset a zoom of the Gly region. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K The residues in the UBA with previous assignments do not shift and are labelled. The peaks which show CSPs upon binding to MAP-LC3 in this region are likely to correspond to the Gly residues in the LIR.

Up to 1 mM unlabelled ubiquitin was added to the ^{15}N-C331S p62 300-440-MAP-LC3 binary complex. More ubiquitin was added than MAP-LC3 because the affinity the p62 UBA has for ubiquitin is known to be weak due to the competitive processes of dimerisation and binding. Again, concentration dependent changes in chemical shift were observed for resonances corresponding to the UBA (figure 8.8). This was indicative of a binding event involving the UBA domain. Saturation of the ^{15}N-C331S p62 300-440 occurred around 1 mM, a 1:4 molar ratio. This was similar to the observed saturation point for the p62 UBA domain in isolation, highlighting that the affinity has not been significantly altered in the longer construct. The addition of ubiquitin resulted in a mixture of slow and fast chemical exchange regimes, again in accordance with the data previously obtained for the p62 UBA and ubiquitin. This would suggest that the mechanism of binding is not affected in the longer p62 construct.

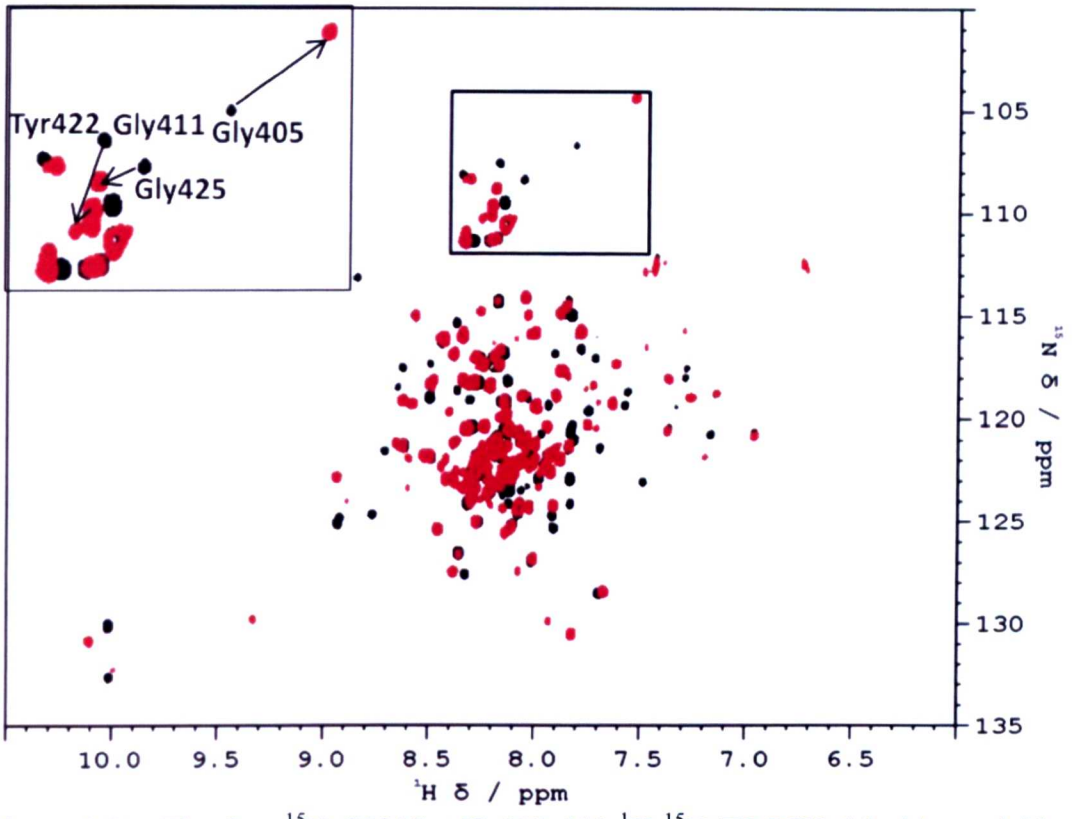


Figure 8.8 – The free ^{15}N-C331S p62 300-440 ^{1}H-^{15}N TROSY (black) overlaid with the ^{1}H-^{15}N TROSY of the ^{15}N-C331S p62 300-440 bound to both MAP-LC3 and ubiquitin (red). Inset a zoom of the Gly region. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K. Significant CSPs are observed for nearly all residues in this region. The peaks corresponding to the UBA show similar CSPs to those previously reported when bound to ubiquitin.

One particularly interesting observation was that some resonances corresponding to the bound form of the UBA in the C331S p62 300-440 construct had deviated from the chemical shifts previously reported for the isolated UBA domain. The resonances corresponding to the UBA only move on the addition of ubiquitin, ruling the possibility of a MAP-LC3-UBA complex. The data therefore suggests a different conformation of the ubiquitin bound UBA in this p62 construct, which could also occur in full length p62 *in vivo*. When examined in detail around half the UBA peaks showed chemical shifts perturbations. Moreover, many could not be unambiguously identified due to their chemical shifts being located within the crowded central region. The conserved MGF motif utilised by the UBA to bind to ubiquitin critically did not show significant CSPs indicating that the mechanism of the interaction with ubiquitin was not affected. The resonances which shifted corresponded to residues in helices 2 and 3 as well as the C-terminus (figure 8.9),

which interestingly have different chemical shifts in the UBA dimer and monomer. This 'alternative bound form' was observed in both titrations highlighting that it was not an artefact of MAP-LC3 binding first. Moreover, because the 'alternative bound form' was observed in the titration where ubiquitin is bound first suggesting that the alternative form is not the result of a direct contact between the MAP-LC3 and ubiquitin.

The C-terminus is likely to adopt a different chemical environment in the longer construct as it is conformationally flexible. However, further investigation into the movement of the peaks in helices 2 and 3 would be required to determine if indeed an 'alternative bound form' exists. This form could provide insights into how the UBA functions as part of the full length p62 protein.

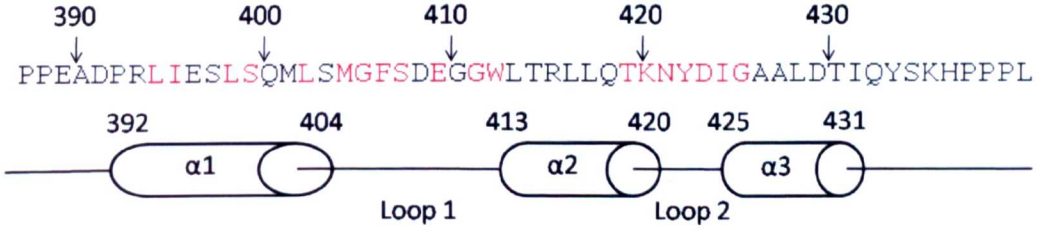


Figure 8.9 –The amino acid sequence of the p62 UBA. The residues in red are assigned and the residues in black not assigned in the bound form. The red residues are the assigned residues of the UBA in the bound form. Helices 2 and 3 and the C-terminal are unable to be assigned. These residues either show different chemical shifts than observed for wild type p62 UBA, or are located in the clustered central region of the spectrum and are unable to be unambiguously assigned.

8.2.2.3 Backbone assignment of the free ^{15}N- C331S p62 300-440

The C331S p62 300-440 construct has a molecular weight of 15.4 kDa. Including the GS extension left from thrombin cleavage, the construct is 143 residues in length. Proteins greater than 10 kDa require high spectral resolution. Triple resonance experiments are required to assign the protein backbone in larger proteins. Separating into a third dimension should reduce overalapping resonances in the ^{1}H-^{15}N TROSY. Therefore a 250 μM ^{13}C^{15}N-C331S p62 300-440 was prepared by dissolving lyophilised protein in 25 mM potassium phosphate, 25 mM NaCl, pH 7. HNCACB, HN(CO)CACB, HNCA, HNCO and HN(CA)CO experiments were set up with the total time not exceeding 48 hours. The HNCA experiment was also recorded to ensure Cα signals were detected as β optimised

data was collected on the HN(CO)CACB. Due to the highly degradable nature of the C331S p62 300-440, the HNCACB and HN(CO)CACB experiments were collected as Band-selective Excitation Short-Transient (BEST) experiments. These experiments shorten the delay between consecutive scans dramatically reducing data acquisition times whilst maintaining the signal to noise ratio in order to collect high resolution data.

However due to the unstructured nature of this construct and the high prevalence of Pro, Glu, Ser and Thr residues of the PEST sequence a full backbone assignment was unlikely. It was therefore decided that assignment of the LIR and UBA only would be the most sensible option. The LIR is only 12 residues in length and it contains a unique and distinctive Ser-di-Gly motif at the start. Both Ser and Gly residues have characteristic Cα and Cβ chemical shifts making them easy to assign using the classical backbone assignment experiments, especially when connected in a sequence. Moreover, the assignments for the UBA had been previously reported and transferring the assignments to the longer construct should not be a difficult task.

Despite the collection of good quality data many of the peaks corresponding to the UBA were not detected by the triple resonance experiments at this concentration. This was not problematic in the free form as the assignments for the UBA were able to be transferred to the C331S p62 300-440. As predicted the assignment of the full length protein would have been challenging. Assigning the Cα, Cβ and CO peaks to individual spin systems were difficult due to the considerable peak overlap which remained despite separating into three dimensions. This prevented the ability to link up distinguishable sequences. The automated assignment panel was also unreliable since it suggests potential residues corresponding to the *i+1* residue and its suggestions were poor.

The assignment process was problematic even when focusing on the 12 residue LIR only. The unique Ser-Gly-Gly sequence was easily identified (figure 8.10) as predicted. However the reminder of the LIR, especially the three consecutive Asp residues, required a different approach as sequential linking of spin systems was not possible. The spin systems which had moved in the ^{1}H-^{15}N TROSYs recorded for the free and MAP-LC3 bound C331S p62 300-440 were noted. These spin

systems were analysed further in the HNCACB, HN(CO)CACB and HNCA experiments. These spin systems were considered to be part of the LIR because they had moved. Using the Cα and Cβ shifts from these experiments the amino acid type for the *i* and *i-1* residues could be deduced. Despite this, only half the LIR was able to be assigned as these were distinctive amino acids (figure 8.11).

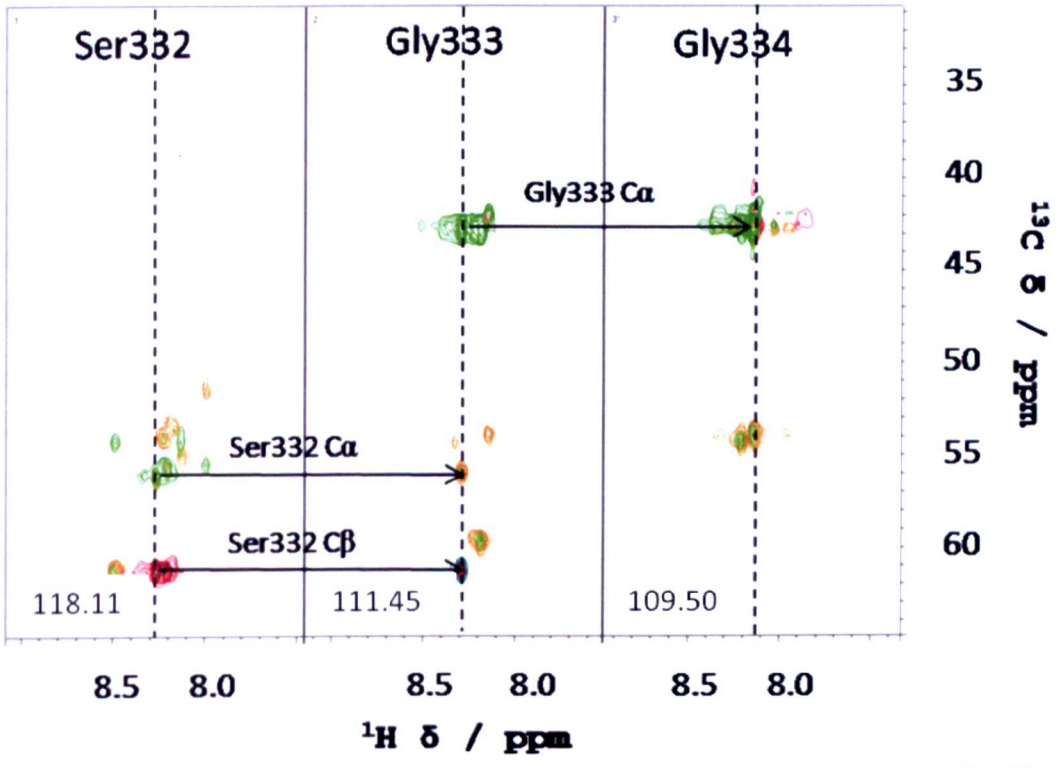


Figure 8.10 – The strips for the Ser-diGly motif in the LIR of the free ^{13}C ^{15}N C331S-p62 300-440. The HNCACB (red, green) and HN(CO)CACB (teal, orange) spectra show the sequential connectivity between Ser332, Gly333 and Gly334. The HNCACB shows Cα and Cβ resonances from residues *i* and *i-1*, whereas the HN(CO)CACB shows only the residues from the *i-1* residue. Each strip shows is mapped to the amide ^{1}H and side chain ^{13}C frequencies, with the ^{15}N amide frequency displayed at the bottom left. The black arrows indicate correlating resonances between connected residues.

There are other distinctive residues present in the LIR: a single Thr and another di-Ser motif. These were logical starting points for this method of assignment. There are two other di-Ser motifs and several other Thr residues in the C331S p62 300-440 sequence. However one of the other di-Ser motif is located centrally in the linker region between the LIR and the UBA it is unlikely to participate in binding and therefore have spin systems which do not move upon MAP-LC3 binding. The other di-Ser motif (located at the start of the LIR) contains Ser332

which was previously assigned as part of the unique Ser-diGly motif. Thr339 is the only Thr residue in the LIR and of the seven Thr residues found in this construct three are found in the UBA and therefore have known chemical shifts, two have been previously identified using the unique PSTG and PTG sequences when attempts were made to assign the p62 341-440 construct[282]. This leaves just two unassigned Thr residues in the C331S p62 300-440 construct. The spin system which was perturbed upon MAP-LC3 binding was likely to be Thr339. To check this, the Cα and Cβ shifts of the preceding residue were checked. Thr339 is located after Trp338. The spin system assigned to Thr339 showed Cα and Cβ which were likely to belong to either a His or Trp residue. The other unassigned Thr residue, Thr305 is preceded by an Ala residue which has a distinctive Cβ shift. Moreover, this residue is located close to the C-terminus which may prevent it from being visualised. Therefore Thr339 was assigned to the moving spin system. Ser331 (the residue preceeding the LIR) was also able to be assigned as it was connected to the Ser di-Gly motif.

The other residues of the LIR which were attempted to be assigned using the method of identifying the amino acid type of *i* and *i-1* residues from perturbed spin systems were Trp338 and Leu341. Since these residues are not distinctive like Ser and Thr residues these assignments were tentative at best. The only residues which were unable to be assigned at all were Asp335, Asp336, Asp337 and His340. If the Asp residues are all located in a fully unstructured region, they would be expected to be in the same environment and therefore possess the same Cα and Cβ shifts prohibiting their identification from one another. This method used to assign the LIR is somewhat unusual. It would be preferable to adopt a classical approach whereby individual spin systems are linked. However, this would require data which did not have significant overlap of resonances.

In the unbound C331S p62 300-440 construct the UBA domain exhibited chemical shifts which were very similar to those observed for the unbound isolated UBA domain. The assignments of 40 resonances were able to be transferred to the ^{1}H-^{15}N TROSY of the C331S p62 300-440. The resonances which were unable to be assigned were located in the crowded central region of the spectrum or were at the C-terminus which exhibits flexibility which can affect chemical shift. The 3D triple resonance spectra were unable to help overcome this

problem as many of the peaks corresponding to the UBA were weak or not detected at all.

VEGATQSLAEQMRKIALESEGRPEEQMESDNSSGGDDDWTHLSSKEV
DPSTGELQSLQMPESEGPSSLDPSQEGPTGLKEAALYPHLPPEADPR
LIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAALDTIQYSKHPPPL

Figure 8.11 – the sequence of the C331S p62 300-440 construct. The LIR and UBA are underlined. The residues which were assigned in the free ^{15}N-C331S p62 300-440 TROSY spectrum are coloured yellow and purple for the LIR and UBA respectively.

8.2.2.4 Backbone assignment of the C331S p62 300-440 when bound to MAP-LC3 and ubiquitin

A 400 μM $^{13}C^{15}N$-C331S p62 300-440 saturated with both MAP-LC3 (400 μM) and ubiquitin (1.6 mM) was prepared by dissolving lyophilised protein in 25 mM potassium phosphate, 25 mM NaCl, pH 7. The concentration of the C331S p62 300-440 construct was increased in order to detect the resonances corresponding to the UBA peaks, which were lost at 250 μM in the triple resonance experiments. This was necessary due to differences being observed between the ubiquitin-bound form for the isolated UBA and the C331S p62 300-440 construct (discussed previously). MAP-LC3 had shown to be insoluble above 250 μM; however, gradually adding lyophilised MAP-LC3 to the $^{13}C^{15}N$-C331S p62 300-440 seemed to prevent this. A ^{1}H-^{15}N TROSY showed resonances which overlaid with a previously recorded spectrum for the MAP-LC3 bound form and confirmed that the MAP-LC3 bound form had been reached before the addition of ubiquitin. The same triple resonance experiments were conducted on the bound sample as the free $^{13}C^{15}N$-C331S p62 300-440, again with the total time for the five experiments not exceeding 48 hours. BEST pulse sequences were used to keep data collection times down. Similar to the free form the data produced from the experiments was difficult to analyse due to significant overlap of resonances, despite separating into three dimensions.

A full assignment of the p62 LIR in the bound form could not be attained due to the high degree of overlap of signals. The residues of the LIR that were assigned confidently in the free form were also able to be assigned in the bound form.

Ser331 (the residue preceding the LIR), Ser332, Ser342 and Ser343 were all able to be tracked to the bound form as they were in fast exchange. Gly333, Gly334 and Thr339 were in slow exchange but were able to be assigned from the triple resonance data using the methodology outlined earlier. This was aided by the distinctive Cα and Cβ shifts for Gly and Thr amino acids. The Ser-diGly motif also showed spins systems which could be linked sequentially (figure 8.12).

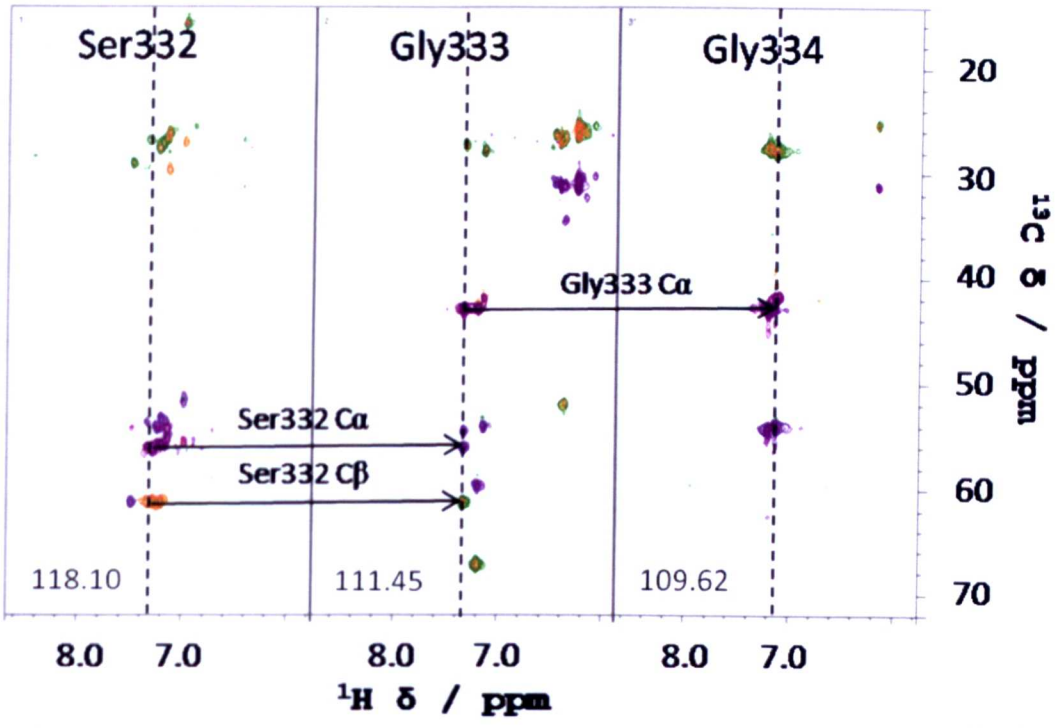


Figure 8.12 – The strips for the Ser-diGly motif in the LIR of the bound ^{13}C ^{15}N C331S-p62 300-440. The HNCACB (deep purple, orange) and HN(CO)CACB (green, bright purple) spectra show the sequential connectivity between Ser332, Gly333 and Gly334. The HNCACB shows Cαand Cβ resonances from residues *i* and *i-1*, whereas the HN(CO)CACB shows only the residues from the *i-1* residue. Each strip shows is mapped to the amide ^{1}H and side chain ^{13}C frequencies, with the ^{15}N amide frequency displayed at the bottom left. The black arrows indicate correlating resonances between connected residues.

As previously mentioned, many of the resonances corresponding to the bound UBA in the ^{15}N-C331S p62 300-440 construct were not found at the same chemical shift in the ^{1}H-^{15}N TROSY as they were for the isolated domain. These residues were clustered in helices 2 and 3 and the C-terminal. Therefore many peaks in the UBA were unable to be assigned in the ubiquitin bound form. Increasing the concentration of the $^{13}C^{15}N$ C331S p62 300-440 sample did enable

the detection of some of the UBA peaks, but not all. Unfortunately the number of scans could not be increased due to the sensitive nature of this protein. Obviously residues which did not produce any resonance peaks in the 3D data could not be assigned. For unassigned residues in the UBA where signals could be detected a different approach was used to attempt assignment. The ^{13}C chemical shift is not as sensitive as the ^{15}N chemical shift. Therefore it is likely that the Cα, Cβ and C' chemical shifts were likely to be similar to those previously reported. Therefore by using the Cα, Cβ and C' chemical shifts which were deposited into the BRMB for the bound form of the UBA attempts were made to assign the missing UBA peaks. However, since the Cα and Cβ shifts are very similar for most types of amino acid this was more complicated than anticipated and consequently none of the missing residues were unable to be unambiguously assigned (figure 8.13).

A

VEGATQSLAEQMRKIALESEGRPEEQMESDNSSGGDDDWTHLSSKEV
DPSTGELQSLQMPESEGPSSLDPSQEGPTGLKEAALYPHLPPEADPR
LIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAALDTIQYSKHPPPL

B

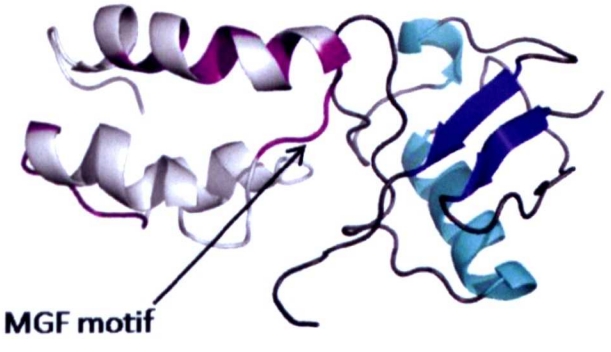


Figure 8.13 – A) the sequence of the C331S p62 300-440 construct. The LIR and UBA are underlined. The residues which were assigned in the bound 15N-C331S p62 300-440 TROSY spectrum are coloured yellow and purple for the LIR and UBA respectively. B) a representation of the p62 UBA monomer complexed to ubiquitin (PDB IDs 1UBQ and 2JY8). Unassigned residues of the bound UBA are coloured grey. Residues which were able to be assigned in the bound UBA are coloured purple. The majority of the UBA is unassigned in the ubiquitin bound form. However, loop 1 which contains the conserved MGF motif which binds to ubiquitin has been assigned highlighting the mechanism of binding has not been affected.

8.2.2.5 Assignment summary

Overall the partial assignment of the LIR and UBA in the bound form precludes the quantitative analysis of the titration data. The distinctive residues of the LIR, namely Ser332, Gly333, Gly334 and Thr339, Ser343 and Ser343 were assigned in both the free and bound forms. However, the only residue from the W-X-X-L motif (which has been previously shown to interact with MAP-LC3)[116,275] which was able to be assigned was Thr339. The inability to assign the perturbed residues observed in the bound form of the UBA prevented any information on the 'alternative bound form' from being gained. Chemical shift perturbation maps, surface plots and accurate K_d values all could not be completed without a full assignment.

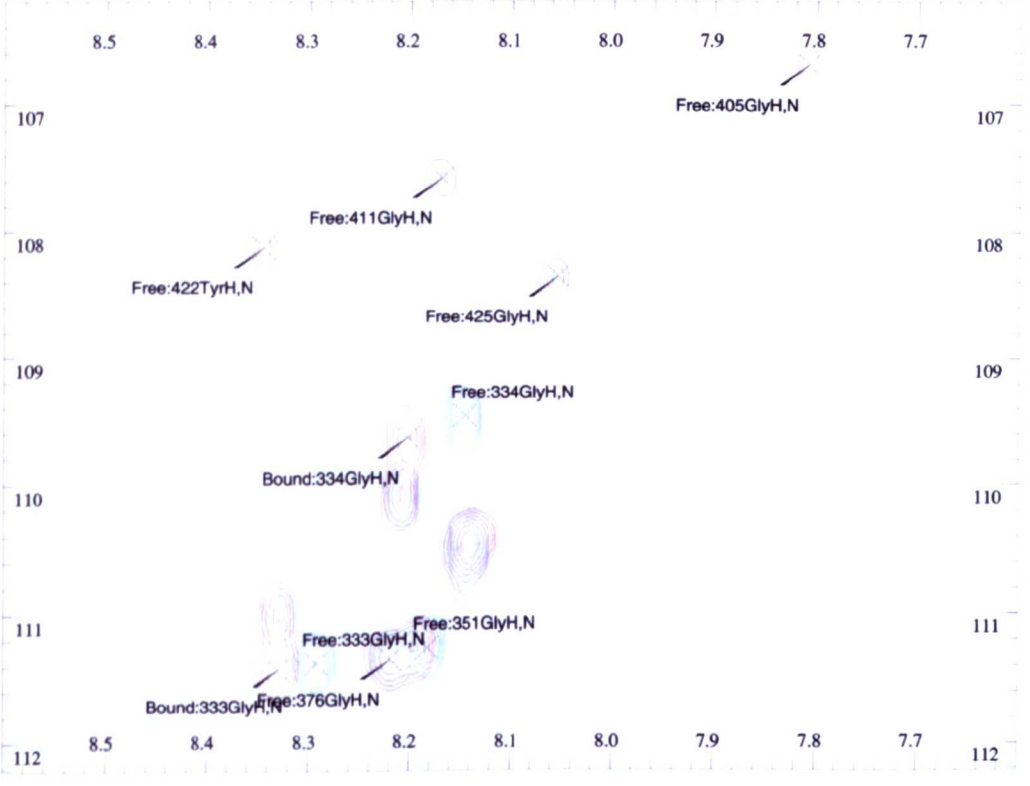


Figure 8.14 – A zoom of the Gly region in the free ^{15}N-C331S p62 300-440 (blue) and the MAP-LC3 bound ^{15}N-C331S p62 300-440 (red). Tyr422, Gly411, Gly425 and Gly405 are part of the UBA. Gly333 and Gly334 were assigned using triple resonance data reported in this chapter. Gly351 and Gly376 were previously assigned by Dr Thomas Garner[282].

8.2.2.6 ^{1}H-^{15}N Heteronuclear NOE experiments highlight which parts of the protein are flexible

During the course of the titrations various ^{1}H-^{15}N heteronuclear NOE spectra were recorded on the ^{15}N-C331S p62 300-440 construct. Experiments whereby the p62 construct was saturated with MAP-LC3, saturated with ubiquitin and saturated with both proteins were recorded (figure 8.15). ^{1}H-^{15}N heteronuclear NOE experiments can reveal flexible regions of the protein. Due to the larger size of the ternary complex, a TROSY based version of the pulse sequences was applied. Two interleaved ^{1}H-^{15}N TROSY experiments, one with proton saturation and one without, were recorded. The spectra are separated and ^{1}H-^{15}N heteronuclear NOE values can be calculated from the ratio of ^{15}N intensities.

The inability to fully assign the LIR and the UBA of the C331S p62 300-440 construct in the free and bound forms precluded the quantitative analysis of the spectra. Moreover the extent of degradation was different for each of the spectra as the data was recorded on different samples at different points in the titrations. This would cause different peak heights and would consequently cause inaccurate NOE values to be calculated. However, the observed resonances are a mixture of positive and negative signals, therefore the regions of the proteins which are flexible can be observed. Highly flexible regions of proteins produce negative signals as they relax at a rate which is faster when compared to the overall tumbling rate of the protein. The negative peaks corresponding to degraded peptides were ignored for this analysis.

When the MAP-LC3 protein is bound some residues in the LIR and other parts of the ^{15}N-C331S p62 300-440 construct still exhibit conformational flexibility as indicated by the negative green peaks (figure 8.15, A). This could reflect the small number of residues involved in MAP-LC3 binding. The peaks corresponding to the UBA all show positive signals as they form an ordered domain. The residues which were able to be assigned to the LIR showed a mixture of positive and negative signals. Peaks corresponding to Gly333 and Gly334 in the LIR are negative and therefore flexible; whereas, the peak corresponding to the side chain of Trp338 in the LIR is positive confirming that this residue is involved directly with the interaction with MAP-LC3. The mixture of signals exhibited by residues in the LIR indicates that only the interacting W-X-X-L motif becomes rigid and

that the other residues remain flexible. The peaks corresponding to the UBA are all positive since they form a folded domain.

In contrast when ubiquitin is bound less of the molecule is flexible (figure 8.15, B). Many of the unassigned peaks in the linker which showed negative signals in the MAP-LC3 bound form are now positive. This is quite surprising as the MAP-LC3 protein is larger than ubiquitin and the accommodation of a larger protein would be expected to affect the flexibility of the complex. However, only a small number of residues are required for interaction the with MAP-LC3 compared to the folded domain required for ubiquitin binding. As expected, peaks corresponding to the UBA domain are all positive upon ubiquitin binding reflecting their highly ordered structure. The assigned peaks the in the LIR exhibit negative signals indicating that are still flexible despite ubiquitin being pre-bound. This is easily visualized by the Gly333, Gly334 and the Trp338 side chain (figure 8.15, A).

Unsurprisingly when both proteins are bound nearly all of the peaks are positive highlighting a ternary complex with limited flexibility (figure 8.15, C). No substantial differences between the ubiquitin bound form and ternary complex were observed. This highlights that ubiquitin binding has more of an influence on the inherent flexibility of the C331S p62 300-440 construct. The assigned LIR residues and the UBA all show positive signals highlighting that these regions become less flexible when bound to their binding partners. The unassigned linker exhibits positive signals suggesting that the C331S p62 300-440 is capable of holding these two proteins in position once they are bound.

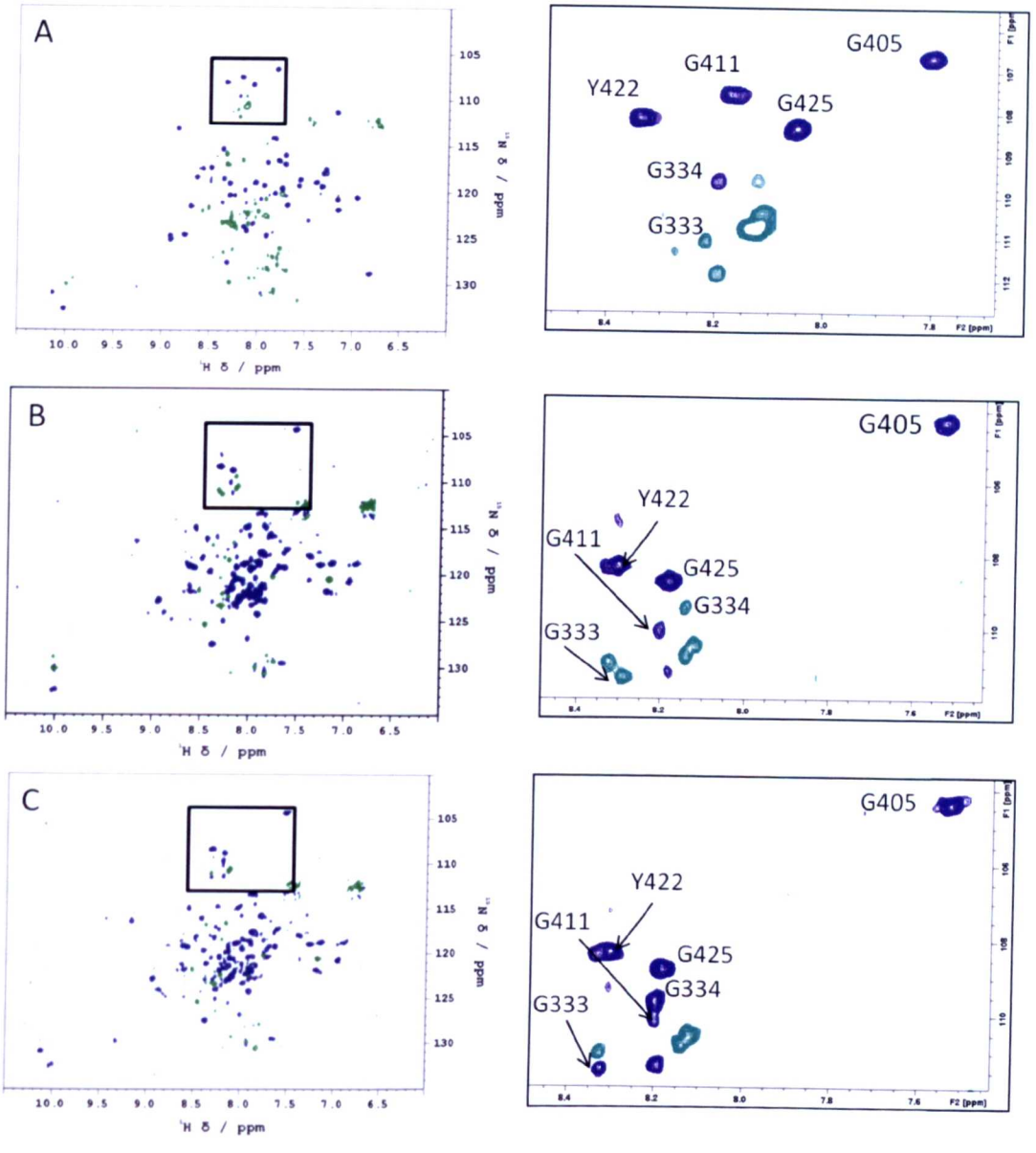


Figure 8.15 – the ^{1}H-^{15}N heteronuclear NOE spectra for the ^{15}N-C331S p62 300-440 combined with A) MAP-LC3, B) ubiquitin and C) both MAP-LC3 and ubiquitin as a ternary complex. The Gly region is zoomed for clarity. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K. The blue and green peaks are positive and negative peaks respectively.

8.2.3 ESI-MS studies on the p62 mediated ternary complex between C331S p62 300-440, MAP-LC3 and ubiquitin

It was clear from the NMR data that a p62 mediated ternary complex was formed between the C331S p62 300-440, ubiquitin and MAP-LC3 proteins. The inability to assign the ^{1}H-^{15}N TROSY prevented quantitative analysis of all of the NMR data. In order to obtain quantitative data, attentions were turned to native ESI-MS. Samples were prepared by dissolving lyophilised protein in 25 mM ammonium acetate.

8.2.3.1 Native ESI-MS of the C331S p62 300-440 construct

A fresh 5 μM sample of the C331S p62 300-440 construct was subjected to native ESI-MS (figure 8.16). Mass ions corresponding to the full length C331S p62 300-440 were detected and a mass of 15481 Da. The mass is in excellent agreement with the predicted mass of 15470 Da. However, the spectrum revealed more degradation than was previously anticipated. Mass ions corresponding to charge states +5 to +13 were visualised, but only the +5 to +8 ions were detected clearly. The higher charge states were surrounded by ions corresponded to the fragmented protein.

The amount of degradation is quite surprising as the NMR data had indicated that the protein was not significantly degraded at the start of an experiment. Although this was not the same sample that was used for the NMR experiments (it is unlabelled), the purification protocol used was the same. The level of degradation would therefore be expected to be similar in each purification. The NMR had revealed that the C331S p62 300-440 was unstable with a short lifespan of approximately 48 hours. ESI-MS was therefore considered to be the ideal technique to use for this sensitive protein as it uses low sample concentrations and has extremely quick data collection times. However, the spectrum produced here indicates the true extent of proteolysis that has occurred during the purification.

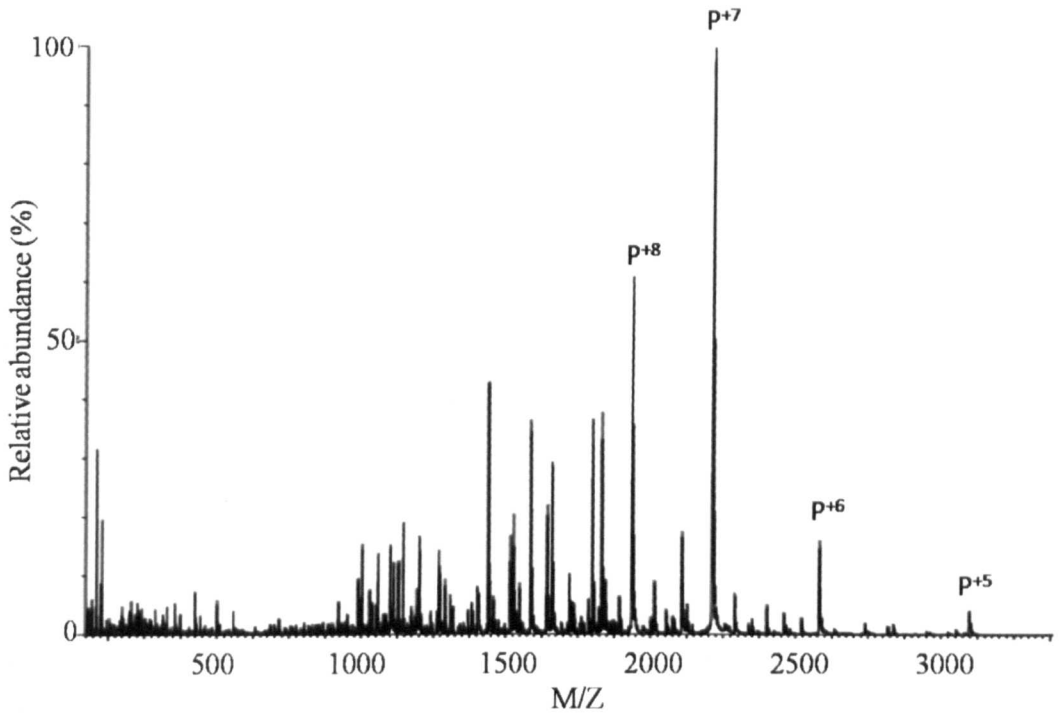


Figure 8.16 – The native ESI-MS spectrum for the C331S p62 300-440 at 5 μM. Spectra were recorded in 25 mM ammonium acetate, pH 7. A spectrum for a fragmented protein is visualised. The +5 to +8 charge states which are clear are labelled although charge states up to +13 were observed in the spectrum.

8.2.3.2 Native ESI-MS of the p62 mediated ternary complex

Despite the significant degradation of the C331S p62 300-440 construct an experiment whereby the three proteins were mixed together at concentrations of 3:1.5:1.5 and 4:2:2 (p62:ubiquitin:MAP-LC3) was conducted. As expected the significant degradation of the C331S p62 300-440 construct prevented a clean spectrum from being produced and any quantitative analysis being completed. However some information can be deduced from the spectrum.

Mass ions corresponding to the three individual free proteins, the two binary complexes and the ternary complex were detected. The spectrum recorded for the 3:1.5:1.5 experiment is shown in figure 8.17. The unbound C331S p62 300-440, ubiquitin and MAP-LC3 proteins recorded masses of 15481, 8563 and 15141 Da respectively and were the dominant species in the spectrum. The 1:1 binary complexes corresponding to the C331S p62 300-440-MAP-LC3 and C331S p62 300-440-ubiquitin, recorded masses of 30624 and 24020 Da respectively. The

binary complexes were detected in vastly different abundances, with much greater intensity observed for the C331S p62 300-440- MAP-LC3 than the C331S p62 300-440-ubiquitin binary complex reflecting the greater affinity p62 has for MAP-LC3 over ubiquitin and correlates well with the NMR data. Mass ions corresponding to a 1:1:1 ternary complex (mass of 39182 Da) was also detected, but in very low abundance. No evidence for a ubiquitin-MAP-LC3 complex was observed highlighting that the two proteins are unlikely to interact with each other once bound to the C331S p62 300-440 construct.

The spectrum recorded in the gas phase is not necessarily a true reflection of the molecules in solution. Complex formation is low in the gas phase, with many complexes detected in low abundance (10-20 %). There are several possible explanations for the low abundance of the ternary complex. The ability to form a protein complex is attributed partly to the role of the solvent, especially in hydrophobic interactions which are extremely weak in the gas phase due to the lack of solvent. The interactions with both MAP-LC3 and ubiquitin are hydrophobic and the combined effects of two significantly weakened interactions are likely to reduce the detection of the ternary complex. The data is consistent with the individual interactions having K_d values in the μM range to produce low populations of binary and ternary complexes. Noncovalent complexes typically dissociate when in flight in the mass spectrometer. This tendency might have resulted in one of the proteins being released from the ternary complex leaving behind the binary complex, or similarly with the two binary complexes the individual proteins would left behind. This would explain the dominance of the individual proteins in the spectrum.

In this particular spectrum it is also possible that binary and ternary complexes are being formed with the different fragments of the C331S p62 300-440 complex. It is clear that the C331S p62 300-440 is the only one of the three proteins to degrade, with the degradation likely to occur at multiple sites in the unstructured region of the protein. It is for this reason that accurate K_d values for the binary and ternary complexes were unable to be calculated from this data.

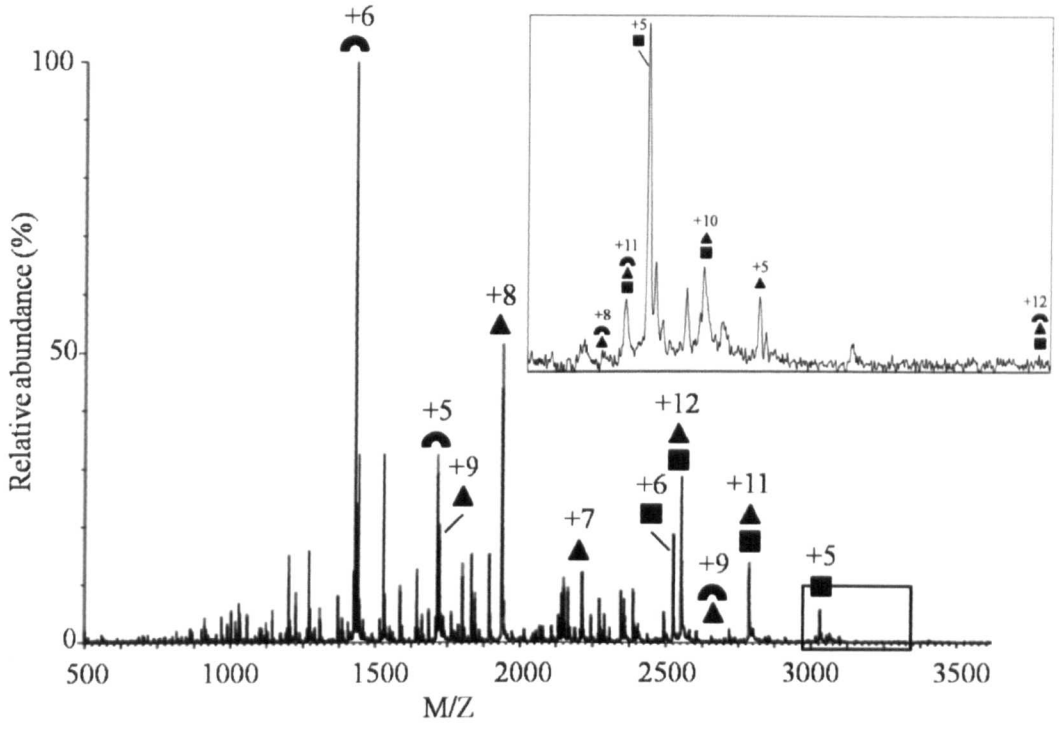


Figure 8.17 – The native ESI-MS spectrum of the C331S p62 300-440, ubiquitin and MAP-LC3 at concentrations of 3:1.5:1.5 respectively. Spectra were recorded in 25 mM ammonium acetate, pH 7. The degradation is the C331S p62 300-440 is clearly visible but the binary complexes and the ternary complex are detected in the spectrum. Inset a zoom of the m/z region 2950-3300 where the ternary complex is found. ▲ is the C331S p62 300-440, ◠ is ubiquitin, ■ is MAP-LC3, ◠▲ is the C331S p62 300-440-ubiquitin complex (1:1), ▲■ is the C331S p62 300-440-MAP-LC3 complex (1:1) and ◠▲■ is the ternary complex (1:1:1).

8.2.4 Modelling the ternary complex

Since sample degradation was a problem in both the NMR and ESI-MS studies, attentions were turned to molecular modelling. The structure of MAP-LC3 bound to the p62 LIR peptide (PDB ID 2K6Q) and ubiquitin bound to the UBA has already been modeled in HADDOCK (PDB IDs 2JY8 and 1UBQ). Residues which form the linker between the LIR and UBA were added to the LIR peptide in the 2K6Q PDB file using Pymol. Residues 387 and 388, both Pros, were added to the linker region to provide a region of overlap. The two PDB's were superimposed and one set of the overlapping Pros deleted to create a single PDB. This file was then energy minimised to produce a model using AMBER (figure 8.18). The energy minimised model revealed a structure that was approximately 0.5 A different to the superimposed PDB file.

The Pro rich PEST domain spans residues 345 – 377 and contains six Pro between the LIR and the UBA. These Pros cause pronounced kinks in the linker region of the chain, which are also capable of restricting the rotation in the linker. Flexible linkers normally allow connecting domains to freely twist and rotate through space to recruit their binding partners. In this case, the high prevalence of Pro residues are likely to prevent an interaction between MAP-LC3 and ubiquitin by restricting rotation whilst maintaining enough flexibility to find their correct binding partners. This is consistent with the results observed by NMR.

Figure 8.18 – A model of the p62 mediated ternary complex built from PDBs 2K6Q, 2JY8 and 1UBQ using AMBER. Residues 332-436 of the p62 protein are depicted in this model. A highly kinked linker region separates the two binding sites on the p62 protein. The kinks are caused by Pro residues which are able to restrict the rotation of the linker, which potentially prevents an interaction between MAP-LC3 and ubiquitin once bound to p62.

8.3 Discussion

The p62 protein's role as a scaffold protein has been clearly established[142]. Interactions have been observed between domains of the p62 or p62 peptides with their binding partners[275,140] but structural analysis of the full length p62 with multiple binding partners is yet to be elucidated. In addition to this, there appears to be a lack of structural information on multiple interactions of other scaffold proteins, such as NBR1, highlighting the need to explore this area further.

It is highly likely that the p62 protein will form various multimeric protein complexes *in vivo*. The p62 has a central role in the regulation of various signalling pathways which are activated in response to different stimuli. Different proteins will therefore be required to bind to the p62 at different times. p62s ability to bind to multiple proteins has however been investigated biochemically. The p62 has been shown to form a ternary complex with aPKC ζ and PAR-4 using coimmunoprecipitation assays. The p62 and PAR-4 proteins are binding partners of the aPKC ζ protein but in this study contacts between PAR-4 and p62 were observed. PAR-4 is inactive when bound to aPKC ζ. Binding of the aPKC ζ to p62 enhances the catalytic activity of the aPKC ζ and interaction with p62 reactivates the inactive PAR-4 protein. The interaction with p62 has a clear role in the regulation of NK-κB activation[283]. In order to gain a greater insight into the molecular recognition events utilised by the p62 with both the PAR-4 and aPKC ζ protein, a detailed structural investigation clearly would need to be conducted.

p62 is capable of interpreting molecular signals by binding to different proteins using distinct binding surfaces. The binding of certain proteins could cause the p62 to become locked in a certain conformation whereby other proteins which regulate other pathways are unable to bind. It is currently unknown if protein recognition sequences found in the unstructured region of the p62 are located spacially close together. For example, the TRAF-6 recognition sequence which occupies residues 228-233 is not sequentially close to the LIR (residues 332-345) but these recognition motifs could be located spacially close together as they are both found in the flexible linker region. Steric clashes could therefore prevent the binding of one protein if another protein has already bound. Alternatively, the flexibility of the linker could be restricted by the presence of a bound protein,

inducing specific locked conformations. These mechanisms could show a mechanism of internal regulation of multiple signalling pathways by the p62.

Until recently, autophagy was thought of as a non-selective degradative pathway. Although the concept of selective autophagy is now generally accepted, very little is known about the molecular mechanisms that govern the specificity. The identification of p62 and NBR1 as adapter proteins in autophagy provided a molecular link between the UPS and autophagy. The idea that p62 is able select certain ubiquitinated proteins and target them to different degradation pathways is attractive. It is possible that the p62 UBA is able to discriminate between different levels of ubiquitination on target proteins, with different linkages and lengths representing distinct molecular signals. Lys48 linked tetraubiquitin has been shown to be the minimum requirement for degradation by the proteasome, it is conceivable that an equivalent exists as signal for degradation by selective autophagy.

With mounting evidence of cross talk between the UPS and autophagy, the human genome is likely to encode several proteins that possess an LIR and some form of UBD. Recently a third protein known as NDP52 was shown to bind to ubiquinated bacteria to target them for autophagic degradation[284]. NPD52, unlike p62 and NBR1, contains a UBZ domain and an LIR. Although research is currently underway into elucidating the exact roles of these adapter proteins in selective autophagy, the role of ubiquitin binding proteins in autophagy in general is only beginning to be unravelled.

Overall, the need to explore the full length p62 protein is evident. Understanding the interplay between the various degradation pathways and the role p62 might have in them is crucial. Many neurodegenerative diseases such as Alzheimer's and Parkinson's disease are caused from the defective action of p62. The p62 could therefore potentially represent a therapeutic target in the future treatment of these diseases.

8.4 Conclusions

The partial degradation of the C331S p62 300-440 construct during the purification has shown to be very problematic during this study. The true extent of the degradation was not revealed until experiments using ESI-MS were undertaken. Moreover, the degradation continued once lyophilised protein had been used to make samples for experiments. The half life of the protein was shown to be around only around 48 hours, which was problematic for 3D triple resonance NMR experiments. The binding studies using both NMR and ESI-MS were therefore only able to be used for qualitative analysis.

However, it was clear from the NMR titration studies that a p62 mediated ternary complex is formed between the C331S p62 300-440 construct, MAP-LC3 and ubiquitin. The binding of the MAP-LC3 protein to the C331S p62 300-440 construct was consistent with the previous study using a peptide corresponding to the p62 LIR and other protein LIR's. The MAP-LC3 interaction in both cases showed a high affinity interaction where saturation was observed around 1:1 molar ratio. Moreover, a mixture of slow and fast exchange was observed by both the p62 and the NBR1 LIR during binding to MAP-LC3 indicating a similar mechanism of binding by both proteins to MAP-LC3. A weak interaction was observed between the p62 UBA and ubiquitin, which again is consistent with the previous data. Saturation occurred around 1:4 (UBA:ubiquitin) molar ratio. Overall the binding of these two proteins to the C331S p62 300-440 is independent of the occupation of other binding site and an allosteric relationship is not observed by the two binding proteins.

The NMR data showed no evidence for contacts between ubiquitin and MAP-LC3 once bound to C331S p62 300-440. CSPs are only observed for the LIR or the UBA upon the addition of MAP-LC3 and ubiquitin respectively. This observation is consistent with the two binding sites functioning as distinct regions on the full length protein. The ^{1}H-^{15}N heteronuclear NOE data highlighted a structure with reduced flexibility for the ternary complex. The limited flexibility of the linker could the prevent associations between the proteins as a 'conformationally locked' molecule could be formed. It would appear from the hetronuclear NOE data that binding to ubiquitin has a greater effect on the flexibility of the linker by restricting free rotation.

The NMR experiments conducted in this chapter are limited to binding to monoubiquitin. It is important to consider differences of p62 binding to longer polyubiquitin chains of different linkages. Longer chains have the potential to prevent access to the LIR, thereby regulating binding by MAP-LC3 and causing an allosteric relationship between the two p62 binding proteins. This is something that should be explored in the future.

Binding experiments using ESI-MS were also able to detect the ternary complex, as well as the two binary complexes. The ternary complex was detected in a 1:1:1 stoichiometry with very low abundance. This is likely to be the result of the individual interactions being hydrophobic. Since the mass spectrometer is bias against hydrophobic interactions it is difficult to maintain and detect intact binary or ternary complex associated by hydrophobic interactions. The two binary complexes were both detected in 1:1 stoichiometries. The intensity of the mass ions corresponding to the C331S p62 300-440-MAP-LC3 binary complex were much greater than the ions corresponding to the C331S p62 300-440-ubiquitin binary complex, reflecting the differences in affinity between the two complexes as both interactions are hydrophobic. Ions for a complex between MAP-LC3 and ubiquitin provided further evidence that the two proteins do not contact each other when bound to p62.

In order to structurally view the ternary complex a model was made based on the existing structures in the PDB. The model showed a linker region with a large number of pronounced kinks corresponding to a high prevalence of Pros. The model showed two structurally distinct binding sites, which were separated by the linker. The Pro residues were likely to restrict the rotation in the linker to prevent association of MAP-LC3 and ubiquitin, whilst maintaining enough flexibility to recruit their correct binding partner.

Overall, the data presented in this chapter demonstrate the difficulties associated with the study of intrinsically disordered proteins. Many proteins possess folded domains separated by unstructured linker regions. The linkers enable the domains to function independently within the protein. Despite the high abundance of intrinsic disorder in modular proteins very little is known about these regions. The lack of information is associated with the limited number of techniques available

which permit the study of these regions; however, NMR is the only technique capable of probing the structure of such regions. As shown by the NMR experiments conducted within this chapter, the chemical shifts of residues within unstructured proteins show a narrow distribution of resonances. Structural investigations into intrsically disordered protein require new NMR experiments to reduce the overlap of resonances. Although, such experiments will require a sample with good stability in order to record high quality data.

8.5 Future work

This project could prove extremely insightful into determining the role of p62 in regulating various degradation pathways. If a homogenous sample of the C331S p62 300-440 construct or indeed the full length p62 protein were able to be obtained several extensions to the project could be made. In order to minimise the degradation a few changes could be made to the purification protocol. All of the purification could be conducted at 4 °C, or extra wash steps could be added to the sepharose column to increase the purity of the thrombin cleaved protein which would therefore remove the need for the gel filtration column. However, in the case of the C331S p62 300-440 construct a methodology of preventing degradation after purification would also be required. This could potentially be achieved via mutagenesis of critical amino acids in the unstructured PEST region.

First of all the free and bound forms of the ^{15}N-C331S p62 300-440 construct would need to be assigned to produce quantitative information from the titration and ^{1}H-^{15}N heteronuclear NOE data. Multi-dimensional NMR has become the leading technique to study intrinsically disordered regions of proteins. X-ray crystallography might be a better technique to examine larger proteins but problems associated with the disordered regions of proteins occur using this technique. Disordered regions make proteins hard to crystallize, and if a crystal is obtained the disordered region is unlikely to be visible on the electron density map.

Since the 3D triple resonance experiments using ^{1}H direct detection for backbone assignment used in this chapter still produced peaks with significant overlap, more complex 3D and potentially even 4D NMR experiments would need to be performed. The development of experiments that exploit ^{13}C direct detection to improve chemical shift dispersion and non uniform sampling methods to reduce experiment times have recently been described (reviewed by Felli *et al*[285]). Moreover, two 4D experiments which also use ^{13}C direct detection known as HCBCACON and HCBCANCO, were used to assign a 12.8 kDa intrinsically disordered protein with a high Pro content. The HCBCACON experiment correlated the resonances of the backbone N of residue *i* with the Hα, Hβ, Cα, Cβ and C' of residue *i-1*; whereas, the HCBCANCO correlates the resonance frequency of C' in residue *i-1* with the Hα, Hβ, Cα, Cβ and N resonances of

residue i^{286}. The obvious limitation to the use of complex NMR experiments is that they are often time consuming and the ^{15}N-C331S p62 300-440 construct has been shown to have a limited life span. Despite the constant improvements in NMR to decrease experiment times it is unlikely that a full set of experiments could be completed on a single $^{13}C^{15}N$ C331S p62 300-440 sample. Potentially a fresh sample would need to be prepared for each individual experiment in order to collect high quality data.

In order to quantify and confirm that the binding of MAP-LC3 and ubiquitin to the p62 is independent from each other binding experiments using ITC could be performed. Although this was observed by NMR in this chapter, accurate K_d values were unable to be obtained. ITC would be able to produce quantitative binding affinities which could be complimentary to NMR. ITC unlike ESI-MS is conducted in solution and would therefore be able to produce more accurate values of K_d. The MAP-LC3 protein could be titrated into C331S p62 300-440 in the presence and absence of ubiquitin, or ubiquitin could be titrated into the C331S p62 300-440 construct in the presence and absence of MAP-LC3. If the affinities produced from the pairs of experiments were comparable the binding is likely to be independent of the occupation of the other binding site. There are some things to consider if these experiments are to work successfully. The MAP-LC3 protein has some solubility problems above 250uM and the p62 UBA dimer adds complexity. The ITC calorimeter is unable to distinguish heat changes associated with the dissociation event and the binding event generating inaccurate binding affinities. Using a UBA mutant which forms a weak dimer such as the T419K mutant (chapter 6) might help to overcome this problem.

A final extension to the project could be to conduct biochemical assays to investigate if proteins with different levels of ubiquitination are targeted to different degradation pathways. Since ubiquitin is used as a signalling molecule it is likely that the fate of the substrate protein is determined by the type of ubiquitination that it becomes modified with. Moreover, the ubiquitin signal could be investigated under proteasome or autophagy inhibited conditions to see if the substrate protein is still able to be degraded or if it is degraded by a different pathway.

9.0 Modelling avidity effects in the binding of oligomeric p62 to polyubiquitin chains

9.1 Introduction

9.1.1 The avidity effect

The avidity effect, also known as the chelate effect, is the co-operative synergistic effects of the individual bonds in multivalent interactions. Avidity effects have been shown to influence the binding of different UBDs to polyubiquitin chains. The presence of two or more UBDs or the attachment of multiple ubiquitins on the same substrate can increase binding affinity through avidity effects and promote ubiquitin-UBD interactions. Avidity might be used by UBDs to filter the noise from non specific transient ubiquitin-UBD interactions and amplify only the output of the desired interaction. For example the Rap80 protein which contains tandem UIM motifs binds to Lys63 polyubiquitin at double stranded DNA breaks (figure 9.1). The UIMs are spacially positioned for avid binding to Lys63 chains, their preferred linkage[287]. In the absence of a binding partner both linear and Lys63 linked chains exhibit flexible structures in which the individual moieties can be regarded as independent and unconstrained subunits.

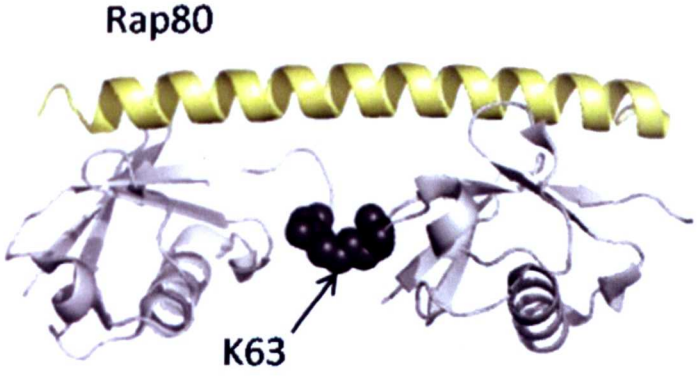


Figure 9.1 – The structure of the tandem Rap80 UIMs in complex with Lys63 diubiquitin (PDB ID 3A1Q).

9.1.2 Oligomeric p62 and the avidity effect

Different UBDs have the ability to discriminate between polyubiquitin chains of different linkage. Linkage preference is conferred via a combination of selectivity and avidity effects to interpret the ubiquitin signal. Avidity effects are thought to have a role in ubiquitin binding by p62, because it is a highly oligomeric protein. p62s ability to self-oligomerise via its N-terminal PB1 domain is clearly evident *in vivo* with p62 found localised in cellular speckles. The individual ubiquitin-UBA interactions are of low affinity; however, avidity effects are postulated to raise the overall affinity when present in oligomeric p62. The p62 protein can therefore present an array of UBAs to long polyubiquitin chains (figure 9.2). However, the recent discovery of the p62 UBA dimer adds complexity to the potential role of avidity in ubiquitin binding by the p62 protein as dimerisation down regulates ubiquitin binding.

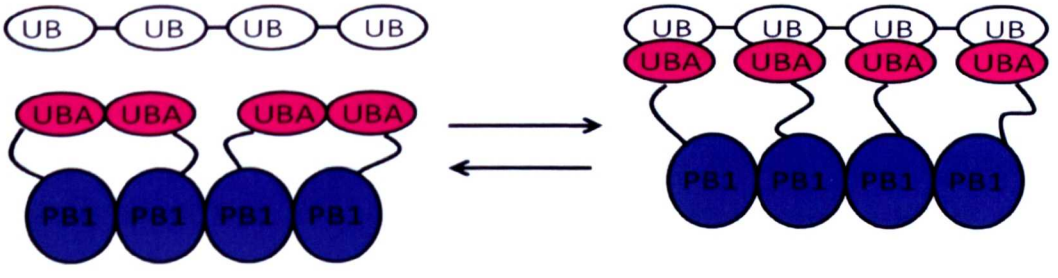


Figure 9.2 – A schematic representation of the full length p62 protein when presented to longer polyubiquitin chains. PB1 oligomers and the UBA dimer are shown in the free and bound forms.

9.1.3 Modelling the full length p62 using GST fusion proteins

To date several studies involving the isolated p62 UBA and its interaction with different ubiquitin ligands have been conducted[139,140,141]. However, GST pull down studies using the full length protein have revealed the p62 has a preference for longer polyubiquitin chains[170,171]. Moreover, differences in affinity were observed for the p62 UBA in the full length protein relative to the UBA domain in isolation[170,171]. The p62 has a high tendency to aggregate, a feature which is mediated by PB1-PB1 interactions. The full length p62 protein was therefore investigated using a model which lacked the PB1 domain and replaced it with an alternative protein oligomer.

The GST protein, present in the pGEX-4T-1 plasmid, forms a highly stable dimer with a K_{dim} of 1 nm. When a GST fusion protein is expressed from this plasmid, it can be used to model the full length p62 protein with the GST component able to mimic the N-terminal PB1 domain (as a dimer). Since the p62 is oligomeric *in vivo* and weak binding had been observed by the p62 UBA to monoubiquitin, it is suspected that avidity effects are used to bind long polyubiquitin chains in order to raise the overall affinity of the interaction. A study highlighted how using GST fusion proteins promoted artificial avidity effects by the dimeric GST bringing together two UBAs in a configuration that favoured simultaneous or avid binding[93]. Oligomeric p62, like the GST fusions, are likely to spatially position UBAs in close proximity to each other. It therefore seemed logical to use GST fusion proteins to model the full length p62 protein. The open conformation of Lys63 linkage tends to be favoured by avid interactions[287], which is consistent with p62s role in NF-κB signalling.

Attempts to purify longer p62 constructs lacking the PB1 domain had previously been conducted by the Searle and Layfield groups, as part of a collaborative project. Therefore some p62 constructs were already cloned and some new constructs were required to be cloned into pGEX-4T-1 plasmid. By expressing different length p62 constructs which extend from the C-terminal as GST fusion proteins, a more realistic model of the full length p62 protein can be created (figure 9.3). When fused to the GST the molecular weights of the constructs increases dramatically to sizes which are at the upper end of the spectrum for proteins which are studied using NMR spectroscopy.

The longer p62 constructs were shown to be susceptible to proteolytic degradation when expressed as thrombin cleaved proteins. Since the GST is a commonly used solubility tag, it was predicted that the GST tag would have a protective role by reducing the susceptibility of the linker region to cleavage.

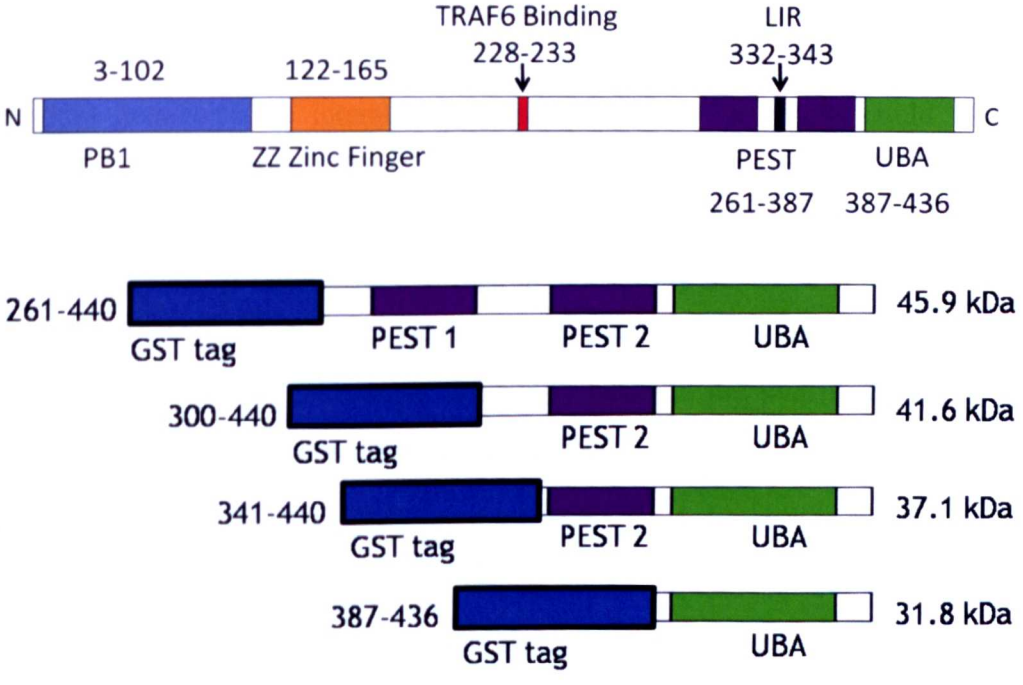


Figure 9.3 – A schematic representation of the GST fusion proteins compared to a schematic representation of the full length p62 protein. The molecular weights of the monomeric GST fusions are indicated to the right of each construct.

9.1.4 Summary and aims

The aim of this investigation was to design a model of the full length p62 protein and investigate the binding affinities associated with monoubiquitin and linear diubiquitin binding. It was postulated that a stronger interaction would be observed in the binding of linear diubiquitin compared to monoubiquitin due to the close positioning of the two UBA monomers, which would favour the avid binding of linear diubiquitin (figure 9.4). Linear diubiquitin was chosen due to its structural similarity to Lys63 diubiquitin[35]. Both linear chains and Lys63 chains have been implicated in regulating the NF-κB signalling pathway, a pathway in which the p62 protein has an integral role[233]. Linear diubiquitin was also easily prepared as it can be cloned sequentially into a plasmid and does not require assembly by ubiquitination enzymes. It is important to note that monoubiquitin from *S.Cerevisiae* and linear diubiquitin from *Homo Sapiens* were used to probe binding. Since ubiquitin is so highly conserved the use of two different ubiquitin orthologues should not affect the results.

NMR spectroscopy and ESI-MS were used to monitor the binding of the GST fusion proteins to monoubiquitin and linear diubiquitin. NMR spectroscopy provides details of the interaction in atomic detail. The results could then be quantified using binding ESI-MS experiments. ESI-MS was considered to be ideal as it was able to be conducted at concentrations low enough to minimise the opposing effects of UBA dimer formation, permitting the calculation of accurate K_d values for the binding interaction. Both NMR and ESI-MS are techniques which could highlight differences in affinity between the two ligands, with any differences likely to be attributed to avidity effects.

The logical approach, given the size of the GST dimer, was to use the smallest construct which consists of the GST tag, the thrombin cleavage site and the UBA, termed GST-p62 UBA, to conduct initial suitability studies. The project could then be extended to studies of the longer constructs in order to build a more realistic model of the full length protein. The avidity effects are predicted to be more pronounced in the shorter constructs, since the effective local UBA concentration is likely to remain high, due to the restricted flexibility in the shorter linker.

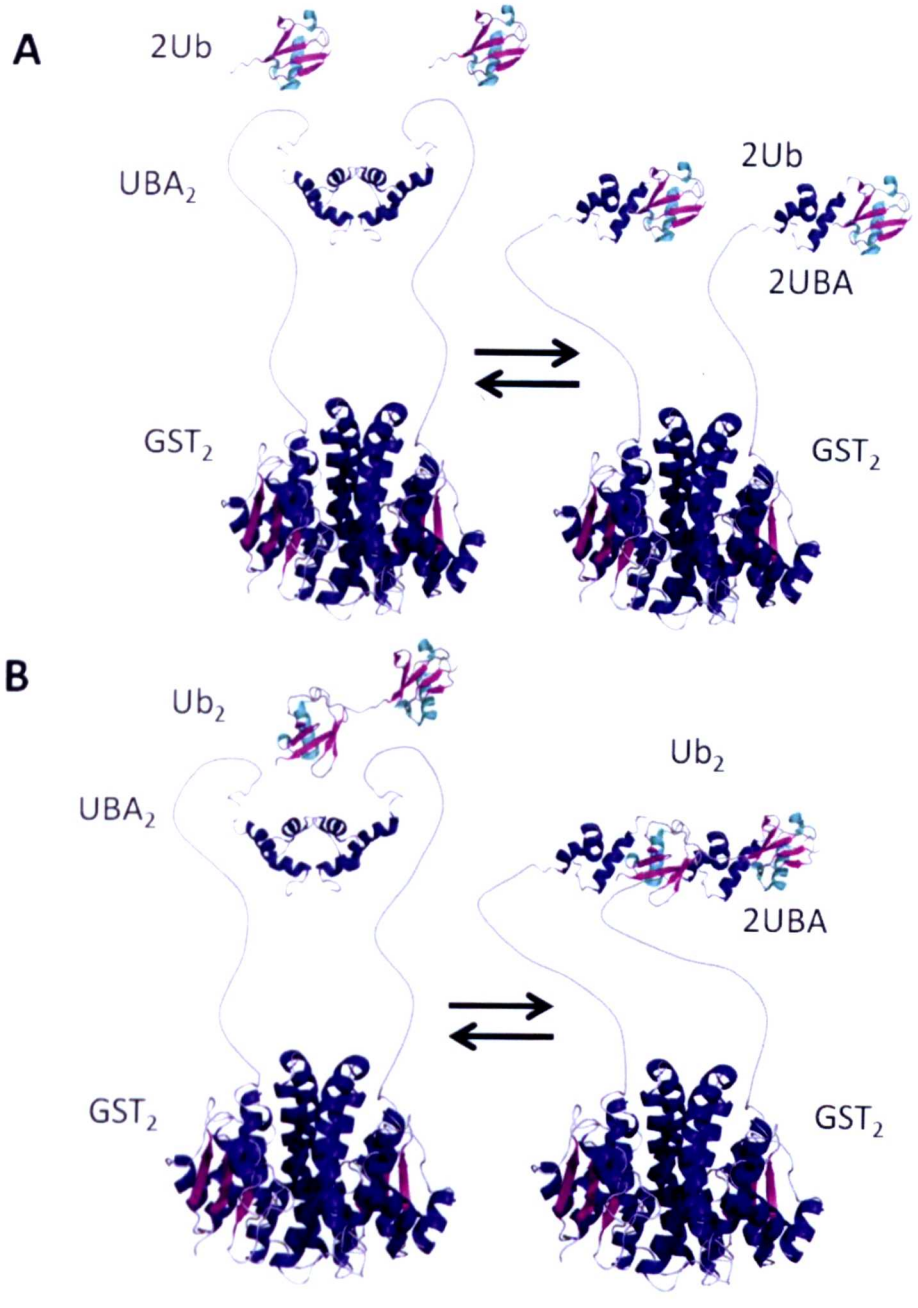


Figure 9.4 – a schematic representation of the GST fusion proteins used to model the full length p62 protein when binding to A) monoubiquitin (Ub) and B) linear diubiquitin (Ub_2). The GST-p62 UBA schematic was made from the structures of the GST dimer (PDB ID 1U87) and either the p62 UBA dimer or monomer (PDB ID's 2KNV and 1Q02 respectively). Predictions as to how monoubiquitin (PDB ID 1UBQ) and linear diubiquitin (2W9N) bind to the GST-p62 UBA are shown. The p62 UBA dimer (UBA_2) dissociates to the biologically active monomer (UBA) in order to bind to the ubiquitin ligand.

9.2 Results

9.2.1 Assessing the viability of the model using NMR on the GST-p62 UBA construct

GST fusion proteins were used in this chapter to model of the full length p62 protein. NMR was chosen as the most desirable technique to monitor the binding of the GST fusion proteins as it provides data in atomic detail. The GST tag used to mimic the PB1 domain, exists as a highly stable dimer with a molecular weight of 52 kDa. Proteins of this size are often unsuitable for NMR studies as they are quite large. Large proteins tumble slowly in solution producing broad NMR signals with low resolution. Small proteins or domains of proteins, such as the p62 UBA, tumble rapidly in solution producing sharp signals by NMR which can be used to determine the structure of the protein and monitor binding interactions upon the addition of a binding partner. In all of the GST fusion proteins there is a linker region separating the GST component from the UBA. The linker enables the GST and the UBA to tumble independently from each other in solution. With this in mind it was predicted that conducting NMR studies using GST fusion proteins would be ideal as the peaks corresponding to the UBA domain would be visible, but those corresponding to the GST, which tumble slowly in solution, would be undetected. The UBA interaction with ubiquitin could therefore be easily monitored at atomic level.

The suitability of this model needed to be assessed prior to any binding experiments. It was decided that the shortest construct consisting of the GST fused to the p62 UBA, termed GST-p62 UBA, would be most appropriate to test the suitability of the model due to its smaller size. In this construct, the C-terminal end of the GST and the thrombin cleavage site (sequence LVPRGS) are unstructured forming a linker between the GST and the UBA domain. A ^{1}H-^{15}N TROSY spectrum was recorded on a 500 μM ^{15}N-GST-p62 UBA sample at 298 K. As predicted, resonances were observed for the p62 UBA residues at the same chemical shifts as those observed for the isolated domain (figure 9.5). This spectrum confirmed that the UBA dimer was also formed in the GST-p62 UBA construct. However, some additional resonances were also observed in the spectrum. These extra resonances were few in number and gave strong signals consistent with the flexible linker residues at the C-terminus of the GST and the six residues of the thrombin cleavage site, although this has not been confirmed

using 3D triple resonance data. Since there were few additional peaks it appeared that the linker, despite being short, was able to separate the GST-p62 UBA construct into two separate domains which are capable of tumbling independently of each other as they would do in a modular protein such as p62. Overall, the spectrum highlighted the GST-p62 UBA protein as a suitable model of the full length p62 protein and that binding studies using this construct should be conducted.

A ^{1}H-^{15}N TROSY spectrum of the GST tag alone was recorded as a control. Many more peaks were visualised in this spectrum than was previously assumed. Despite the GST protein not being entirely NMR silent, it is clear that when fused to the p62 UBA domain, nearly all of the sharp resonances correspond to the UBA. The linker, despite being short, has enabled the two components to tumble independently of each other, with the signals for the GST broadened to undetectable levels.

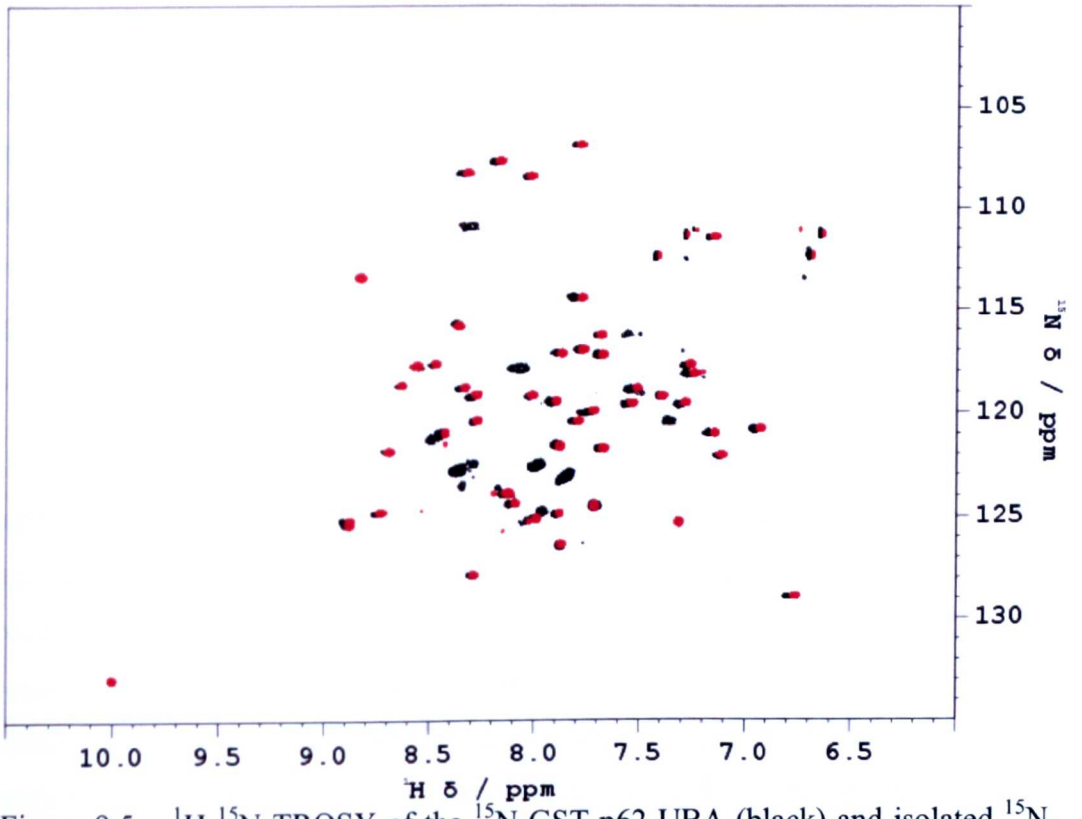


Figure 9.5 – ^{1}H-^{15}N TROSY of the ^{15}N-GST-p62 UBA (black) and isolated ^{15}N-p62-UBA (red). Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

9.2.2 Monitoring avidity effects by the GST-p62 UBA using NMR spectroscopy

The preliminary NMR data using the ^{15}N-GST-p62 UBA as a model for the full length p62 protein showed encouraging results. The GST component is largely invisible by NMR meaning that the UBA can be clearly monitored in binding studies with different ubiquitin ligands. The 'NMR silent' GST would allow the assignments which had been previously determined for the UBA to be transferred to the GST-p62 UBA construct. NMR titration studies whereby monoubiquitin and linear diubiquitin were added to ^{15}N-GST-p62 UBA were completed in order to monitor any avidity effects that might occur. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K

9.2.2.1 Forwards titrations

The addition of up to 2 mM unlabelled monoubiquitin or 1mM unlabelled linear diubiquitin was added to a 500 μM ^{15}N-GST-p62 UBA sample (as separate experiments). Concentration dependent changes in chemical shifts were observed in both titrations indicative of a binding to both monoubiquitin and linear diubiquitin (figure 9.6). The changes in chemical shift were consistent with each other showing the same bound form for the two different ubiquitin ligands. Moreover, the free and fully bound forms of the ^{15}N-GST-p62 UBA showed near identical chemical shifts to the free and bound forms of the isolated ^{15}N-p62 UBA enabling the assignments to be transferred from the ^{15}N-p62 UBA to the ^{15}N-GST-p62 UBA data. Since the chemical shifts for the bound form were so similar to the isolated domain it is unlikely that the GST also interacts with ubiquitin.

At 250 μM the free monomer peaks were not visible meaning that only peaks predominantly in slow exchange were visualised. For resonances in slow exchange, the intensities of the peaks were concentration dependent consistent with a monomer-dimer equilibrium. Some resonances, however, did exhibit fast exchange. These were resonances where the chemical shift for the dimer and monomer were the same and the residue was involved in binding. This is in excellent agreement with the previous data recorded for the isolated UBA. A mixture of slow and fast exchange regimes combined with the near identical chemical shifts for the free and bound forms highlighted a common mechanism of binding between the p62 UBA and the GST-p62 UBA, whereby the UBA dimer must dissociate in order to bind to ubiquitin.

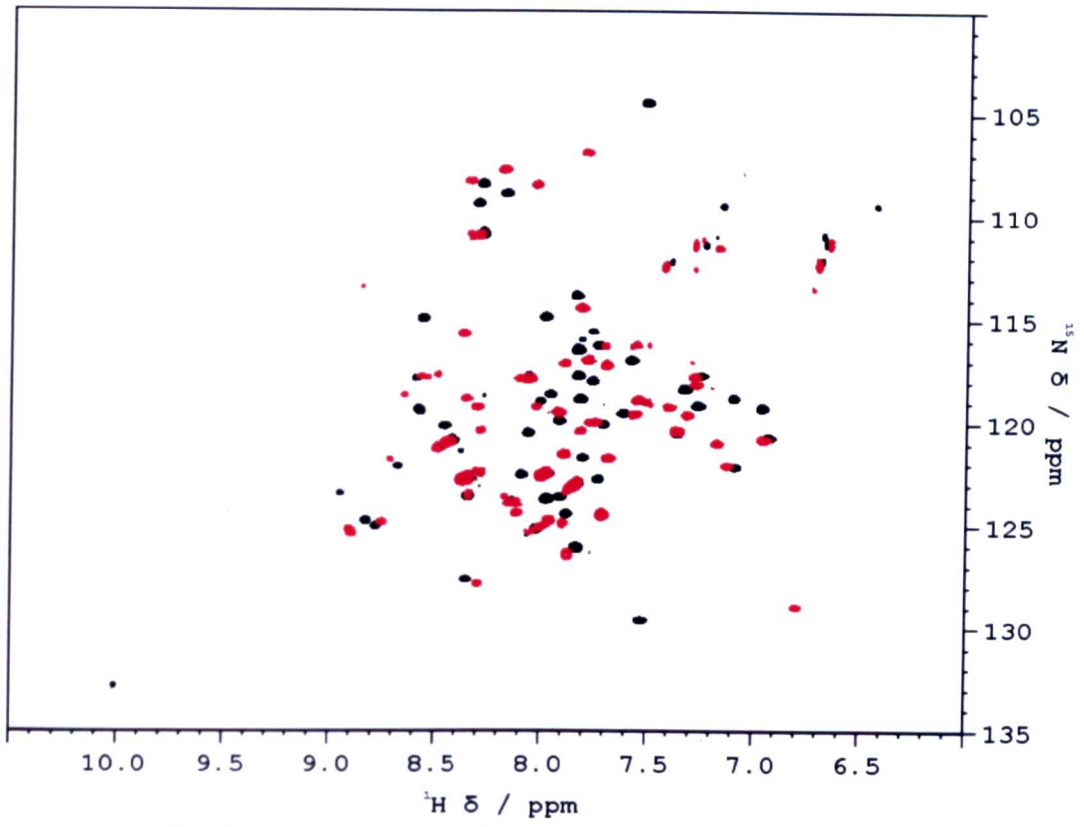


Figure 9.6 – ^{1}H-^{15}N TROSY of the ^{15}N-GST-p62 UBA in the free form (red) and fully bound to monoubiquitin form (black). The same distribution of chemical shifts was observed for the bound form when bound to linear diubiquitin indicating the same mechanism of binding for the two ubiquitin ligands. Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

By plotting the CSPs for each residue between the free and bound forms, regions of the protein which are involved in ubiquitin binding can easily be visualised. The residues involved in binding are largely confined to the end of helix 1, loop 1 and helix 3, with the overall CSP pattern for binding to monoubiquitin and linear diubiquitin showing near identical values (figure 9.7). The only difference being that Gly410 was able to be assigned all the way through the titration with linear diubiquitin. The consistency in the CSP values for binding to monoubiquitin and linear diubiquitin indicates that it is unlikely there is a secondary binding site on the UBA in the presence of linear diubiquitin. Since linear diubiquitin adopts an open conformation, recognition of single ubiquitin moiety in a "beads on a string" mechanism is likely to occur. The overall CSP pattern is very similar to the CSP pattern by the isolated p62 UBA and monoubiquitin with values which were

comparable between the three datasets. The high correlation between the datasets further confirms that the GST does not interfere with the binding interaction.

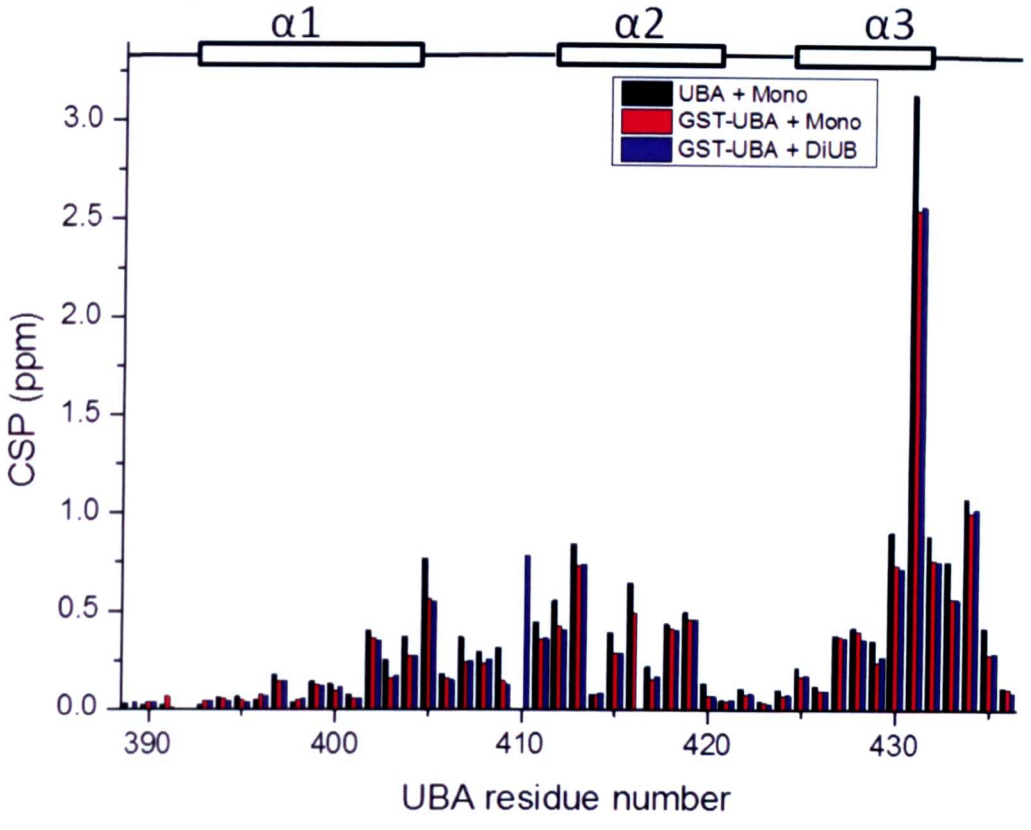


Figure 9.7 – Reside specific CSP for the ^{15}N-GST-p62 UBA and ^{15}N-p62 UBA. Although the CSP values are not exactly the same the general trend correlates well between the two datasets.

However some differences were observed between the ^{15}N-GST-p62 UBA construct and the isolated domain as the ^{15}N-GST-p62 UBA construct showed a higher affinity interaction. The ^{15}N-GST-p62 UBA was saturated at 500 μM monoubiquitin and 250 μM linear diubiquitin, corresponding to a 1:1 and 1:0.5 (UBA:ubiquitin) molar ratio respectively. The faster saturation is consistent with there being two binding sites in linear diubiquitin as it takes twice the amount of monoubiquitin to saturate the ^{15}N-GST-p62 UBA compared to linear diubiquitin. The isolated domain showed saturation of a 1:4 molar ratio. It is possible that the increased affinity observed by the ^{15}N-GST-p62 UBA construct is the result of a weaker dimer being formed by the UBA (as shown in chapter 6). The linker region is potentially too short to permit complete flexibility of the two UBA domains preventing the formation of a high affinity dimer.

At high concentrations the effects of both dimer dissociation and ubiquitin binding of the p62 UBA are observed, therefore an accurate K_d for binding is unable to be determined. However, by plotting binding isotherms for non-overlapping residues in the ^{1}H-^{15}N TROSY spectrum, knowledge on the affinity of an interaction can be gained. Avidity effects are observed by the ^{15}N-GST-p62 UBA when binding to linear diubiquitin as shown by the steeper gradient in figure 9.8. The GST therefore positions the two UBA monomers in close proximity to one another promoting the formation of the UBA dimer. The UBA dimer then favours avid binding to linear diubiquitin compared to monoubiquitin. The increase in affinity for the GST-p62 UBA is not as substantial for linear diubiquitin compared to monoubiquitin.

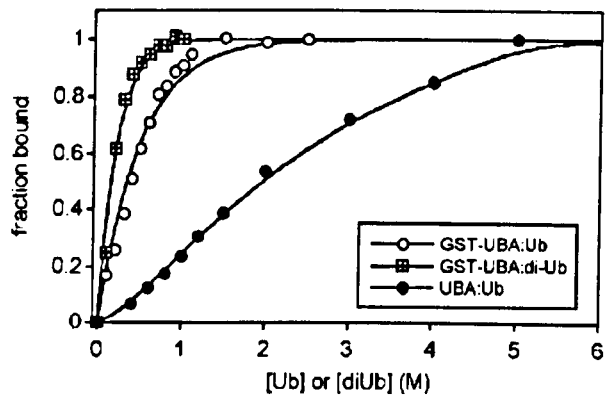


Figure 9.8 – binding isotherms for the isolated p62 UBA and monoubiquitin, the GST-UBA and monoubiquitin and the GST-UBA and linear diubiquitin. The curves represent the data averaged for 26 residues.

9.2.2.2 Identifying the binding surface on ubiquitin

Complementary titrations whereby ^{15}N-monoubiquitin and ^{15}N-linear diubiquitin were titrated into unlabelled GST-p62 UBA (as separate titrations) were also completed. These titrations would identify if the canonical Ile44/Val70 patch is still utilised by ubiquitin in binding, despite the UBA existing as part of a GST-fusion protein. Since the GST-p62 UBA is in excess in these titrations and it is by far the bigger molecule it would require a large amount of GST-p62 UBA to reach a 1:4 molar ratio. Therefore a lower molar ratio would be required in order to keep the GST-p62 UBA soluble in a small volume (600 µl). A 1:2 molar ratio seemed a reasonable comprimise given that the forwards titrations showed saturation of 1:1 or 1:0.5 for the monoubiquitin and linear diubuiqitin respectively.

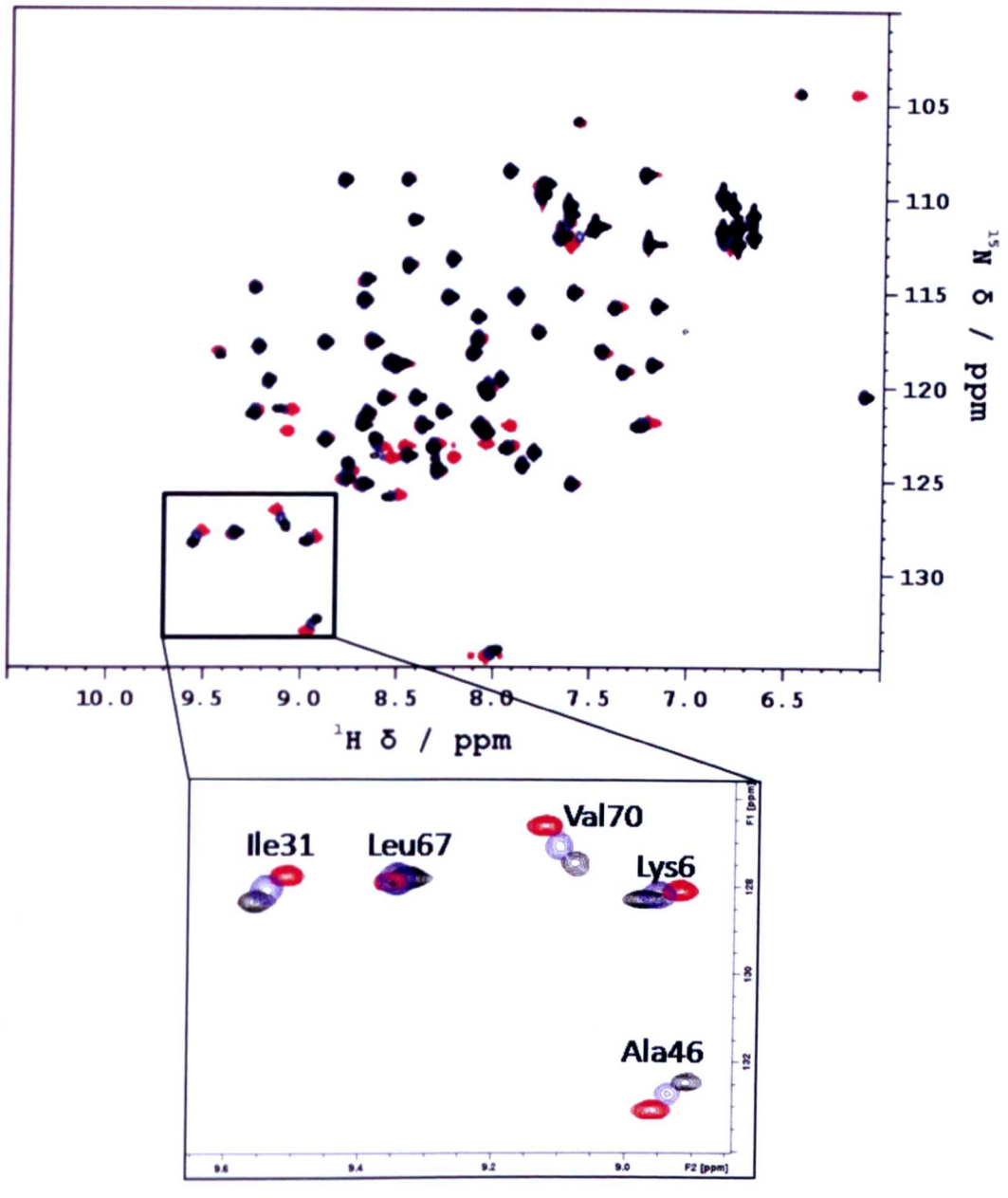


Figure 9.9 – ^{1}H-^{15}N TROSY spectrum for ^{15}N-monoubiquitin (yeast) in the presence of 0 μM (red), 250 μM (blue) and 500 μM (black) GST-p62 UBA. Inset a zoom of some of the residues which perturb upon binding. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

Up to 500 μM unlabelled GST-p62 UBA was titrated into 250 μM ^{15}N-monoubiquitin (figure 9.9). Concentration dependent changes in chemical shift were observed consistent with a binding event. Resonances were observed for 72 out of 76 residues. The missing residues corresponded to the two Pro residues in yeast ubiquitin, Met1 and Gly53. Fast exchange was observed for nearly all residues. However, some peaks broadened during the titration and are therefore in intermediate exchange. Intermediate exchange was observed for Arg42, Ile44, Lys48, Leu69 and Leu71. All of these resonances broaden significantly until they can no longer be detected and never return within the concentration range of this experiment. Arg42, Lys48 and Leu71 broaden very early in the titration, similar to the reverse titration between ^{15}N-monoubiquitin and the isolated p62 UBA. Arg42 had broadened to an undetectable level after the addition of 75 μM GST-p62 UBA; whereas, Lys48 and Leu71 had disappeared by the addition of 100μM GST-p62 UBA. Ile44 and Leu69 broaden later in the titration, after the addition of 175 μM and 200 μM GST-p62 UBA respectively. All residues which exhibit intermediate exchange kinetics are clustered in or around the canonical Ile44/Val70 binding patch. The additional resonances showing intermediate chemical exchange kinetics is indicative of slower on and off rates for complex formation and dissociation compared to the interaction between the isolated p62 UBA domain and monoubiquitin. Since fast exchange is often associated with weak interactions, the presence of a greater number of resonances in intermediate exchange represents a stronger interaction.

A separate titration was completed by titrating up to 500 μM GST-p62 UBA into 250 μM ^{15}N-linear diubiquitin (figure 9.11). Assuming the two subunits of the diubiquitin moiety are in identical environments resonances in each ubiquitin moiety will be present at identical chemical shifts. Therefore a spectrum for monoubiquitin only will be visualised. However many of the peaks had split into two, highlighting that the two ubiquitin moieties are not in identical environments (9.10). The regions expected to differ in each of the ubiquitin subunits are the terminal regions as they are either exposed or connected to the other ubiquitin. When plotted to the surface of ubiquitin peak splitting was indeed observed for residues close to the terminals. The extent of peak splitting varied between residues. Some showed two separate peaks next to each other, whereas others

such as Ser65 and Ile61, only showed partial splitting. For residues which had split into two, only one of the peaks was assigned. The observed spectrum for ^{15}N-linear diubiquitin was slightly different to the spectrum for the free ^{15}N-monoubiquitin because two different ubiquitin orthologues were used.

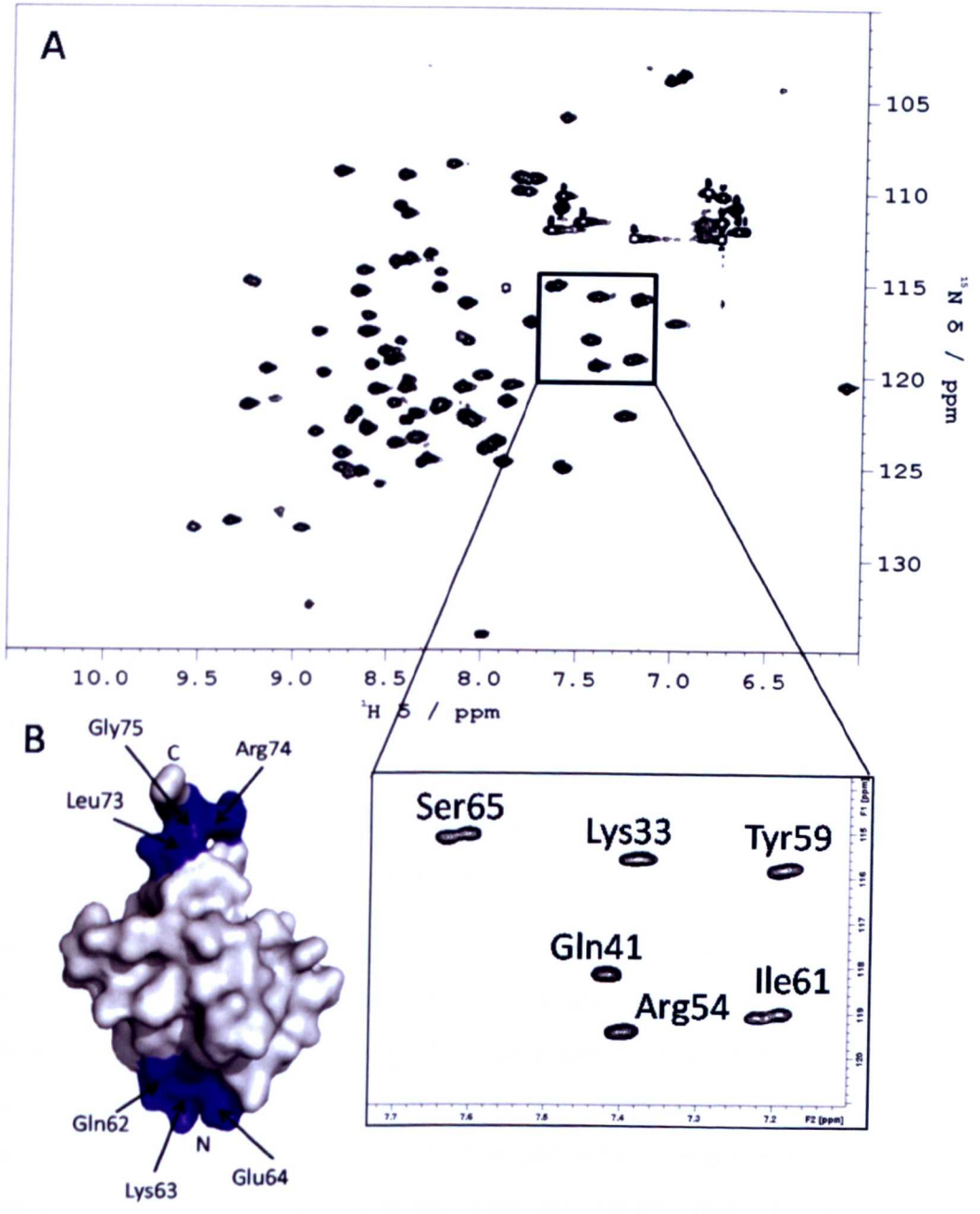


Figure 9.10 – A) The ^{1}H-^{15}N HSQC of the free form of linear diubiquitin. Inset a zoom of residues Ser65 and Ile61 which have split into two due do slightly different chemical environments. These residues are located close to the C-terminals of each of the ubiquitin subunits. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K. B) The surface plot of ubiquitin (1UBQ) with residues with peaks that had split into two highlighted in blue.

Resonances were observed for 70 out of 76 residues. The missing residues correspond to the three Pro residues in human ubiquitin, Met1, Gly53 and Glu24. A mixture of fast and intermediate exchange was observed for residues, although predominantly fast exchange was exhibited. Analogous to the titration between ^{15}N-monoubiquitin and the GST-p62 UBA peaks corresponding to Arg42, Ile44, Lys48, Leu69, Leu71 all broaden to undetectable levels during this titration. Arg42, Lys48 and Leu71 also demonstrate this during the titration between ^{15}N-ubiquitin and the isolated p62 UBA. Arg42, Lys48 and Ile44 are no longer at measurable intensity after the addition of 50 μM, 75 μM and 150 μM GST-p62 UBA respectively. Leu69, Leu71 and Arg72 are lost much later in the titration, after the addition of 250 μM GST-p62 UBA. Arg72 exhibited fast exchange in the titrations between ^{15}N-monoubiquitin and the GST-p62 UBA and ^{15}N-monoubiquitin and the isolated p62 UBA. However when titrated with linear diubiquitin Arg72 changes to intermediate exchange. As well as an additional peak in intermediate exchange, several peaks which are in fast exchange broaden significantly but stay at measurable intensity during this titration are also observed. These peaks correspond to residues Leu8, Ile13, Ala46, Gly47, Leu50 and Val70 and are found either in or around the canonical Ile44/Val70 binding patch, forming a larger region of the molecule which is affected by binding. Overall this indicates a stronger binding interaction occurs for linear diubiquitin when bound to the GST-p62 UBA construct when compared to the interaction between monoubiquitin and the GST-p62 UBA.

In both of the reverse titrations saturation was not reached. This was confusing given the rapid saturation points observed in the forwards titrations. A higher excess of the GST was not attempted to reach saturation as the protein was likely to precipitate out of solution. Saturation was likely to be restricted by the competing equilibria in the GST-p62 UBA, although the mechanisms behind this remain unknown as reverse titrations using the isolated p62 UBA and various mutants have all managed to reach sautration. Since binding to the GST-p62 UBA showed more residues which were in intermediate exchange, it is possible that these residues would return to measurable intensity at saturation. Since saturation was not reached accurate K_d values could not be determined for the two interactions.

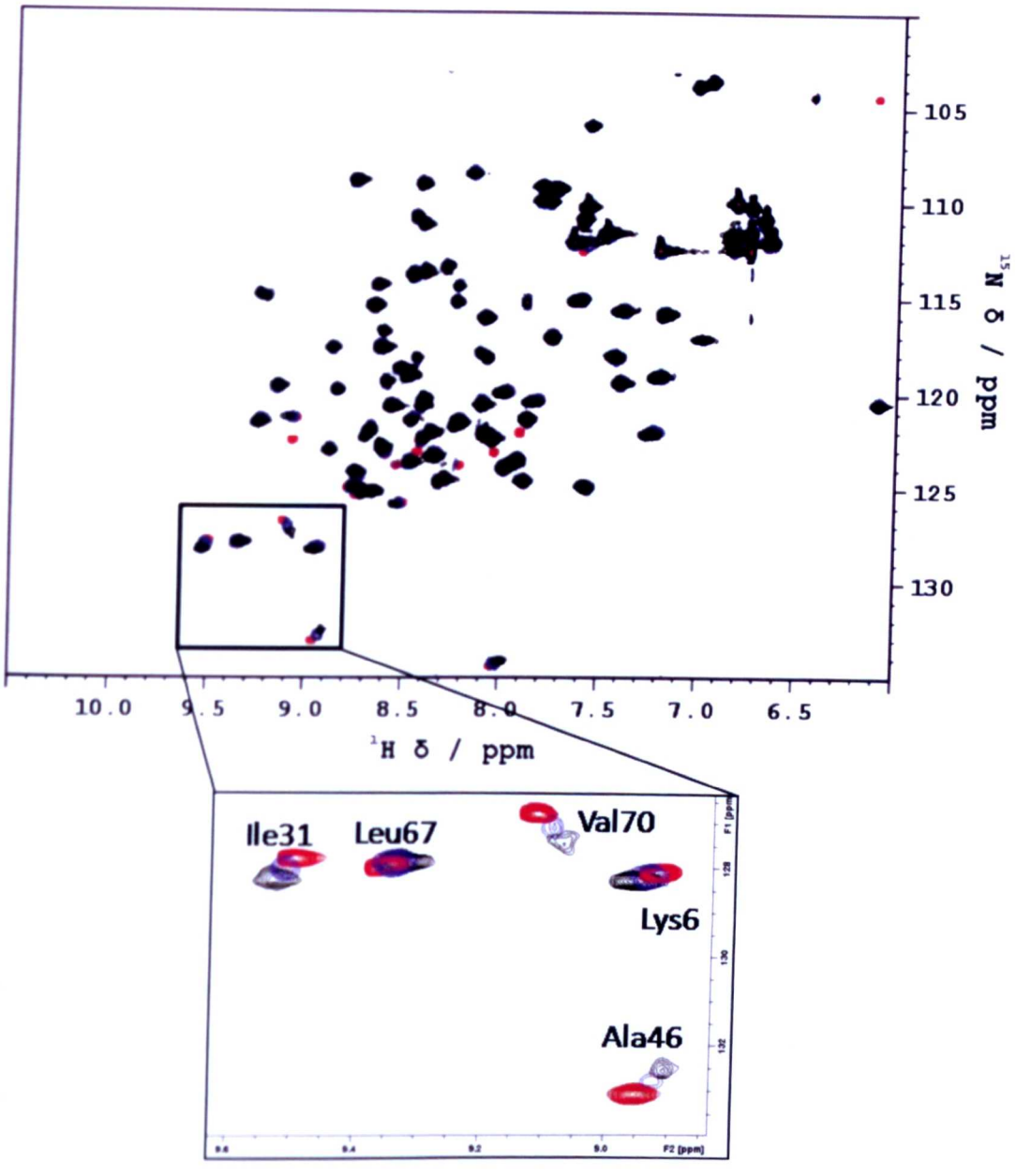


Figure 9.11 - ^{1}H-^{15}N TROSY spectrum for ^{15}N-linear diubiquitin (human) in the presence of 0 μM (red), 250 μM (blue) and 500 μM (black) GST-p62 UBA. Inset a zoom of some of the residues which perturb upon binding. Residues Val70 and Ala46 are examples of the residues which broaden significantly during the titration. The spectrum was recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

The CSPs were plotted for each of the residues in the two titrations (figure 9.12, A). The CSP maps identified clusters of large CSPs corresponding to a previously observed binding epitope on ubiquitin, centred on residues Leu8, Ile44 and Val70. This signifies that the canonical Ile44/Val70 binding patch is utilised by both monoubiquitin and linear diubiquitin to bind to the GST-p62 UBA. Smaller CSPs were also observed relative to the titration between ubiquitin and the isolated p62 UBA domain. Smaller CSP values also indicate that saturation had not been reached during the titrations. Generally, slightly larger CSPs were observed by monoubiquitin when compared to linear diubiquitin. This indicates that avidity effects are responsible for the increase in affinity, as the sum of the two interactions in diubiquitin raise the overall affinity of the interaction.

Residues with significant CSPs (> 0.05 ppm) were plotted onto the surface of monoubiquitin (PDB ID 1UBQ) in order to visualise the region of the protein used to bind to the GST-p62 UBA construct (figure 9.12, B and C). The residues which were in intermediate exchange were coloured in black; whereas, residues in fast exchange were coloured in red. The larger CSPs were represented by darker red colours. A larger region of the molecule is shown to be affected for linear diubiquitin indicative of a tighter binding interface. The presence of an additional resonance in intermediate exchange and several residues surrounding the binding patch which broaden to weak intensities confirm that a larger region is affected, although only Arg42 was plotted onto the surface of the molecule.

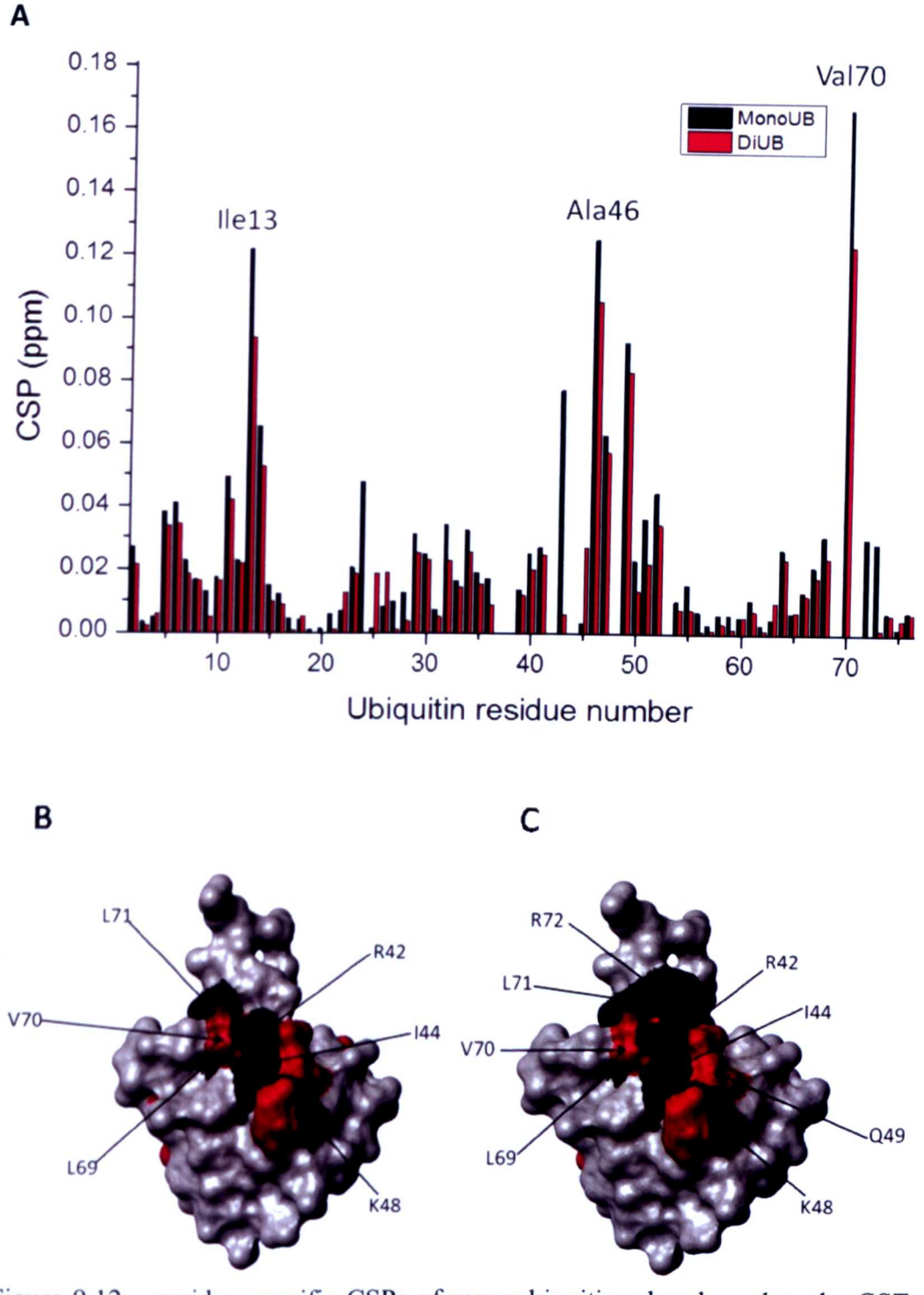


Figure 9.12 – residue specific CSPs of monoubiquitin when bound to the GST-p62 UBA (blue) and linear diubiquitin bound to the GST-p62 UBA (red). The strong correlation in the data highlights that the canonical Ile44/Val70 binding patch is being utilised to bind the GST-p62 UBA. CSPs plotted onto the surface of ubiquitin (PDB ID 1UBQ) for B) monoubiquitin and C) linear diubiquitin when bound to the GST-p62 UBA. Residues Arg42, Ile44, Lys48, Leu69 and Leu71, which broaden to unobservable levels during the two titrations are coloured black. Arg72 also exhibits intermediate exchange in linear diubiquitin, so is also coloured black in C. CSPs greater than 0.05 ppm are coloured red. The darker the red, the larger the CSP. A larger interaction surface is present for the interaction with linear diubiquitin and GST-p62 UBA.

9.2.3 Titrations using hHR23A UBA2 as a control

In order to confirm that avid binding effects are observed by the dimeric UBA of the GST-p62 UBA construct, control titrations using a known monomeric UBA expressed as a GST fusion protein were also conducted. The hHR23A UBA2 domain was the UBA chosen to act as a control UBA. This UBA has been previously shown to be structurally very similar to the p62 UBA with no previous evidence that it exists as a dimer[95,96]. Again all spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

9.2.3.1 Low concentration NMR of the hHR23A UBA2 domain

The discovery of the UBA dimer in the p62 was initially revealed by low concentration NMR. In order to confirm that the isolated hHR23A UBA2 exists as a monomer, a low concentration (10 μM) ^{1}H-^{15}N HSQC was completed. The resonances visualised in this spectrum were subsequently compared to the resonances observed in a high concentration (500 μM) ^{1}H-^{15}N HSQC. At low concentration the number of peaks in the spectrum doubles for the isolated p62 UBA, with the extra peaks corresponding to a low population of the p62 UBA monomer. This was not observed for the low concentration spectrum for the hHR23A UBA2. Peaks at identical chemical shifts but with weaker intensity were observed at low concentration (figure 9.13), thus confirming the presence of a monomeric species only. This is consistent with all the previous structural data on the hHR23A UBA2[95,96].

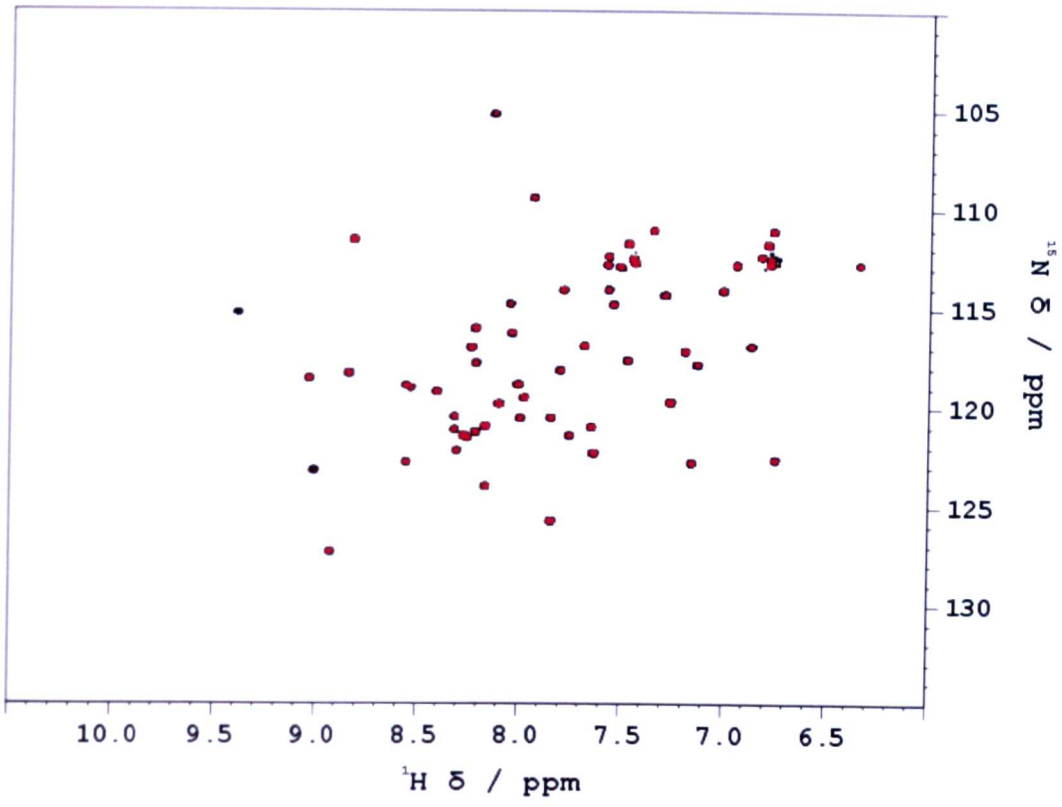


Figure 9.13 – ^{1}H-^{15}N HSQC spectrum of the hHR23A UBA2 at 500 μM (black) and 10 μM (red). Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

9.2.3.2 GST-hHR23A UBA2 compared to hHR23A UBA

Similar to the ^{15}N-GST-p62 UBA, the ^{15}N-GST-hHR23A UBA2 produced a ^{1}H-^{15}N TROSY spectrum with peaks predominantly corresponding to the UBA domain (figure 9.14). Some additional resonances were also observed highlighting the GST as not being completely NMR silent in this construct either. Again these peaks are believed to correspond to the unstructured C-terminal of the GST and thrombin cleavage site which form the linker between the GST and the UBA. In addition to this, the resonances in the spectrum for the free ^{15}N-GST-hHR23A UBA2 produced slightly broader signals than the spectrum for the ^{15}N-GST-p62 UBA, consistent with the molecule showing increased flexibility due to the UBAs not being tethered together as a dimer.

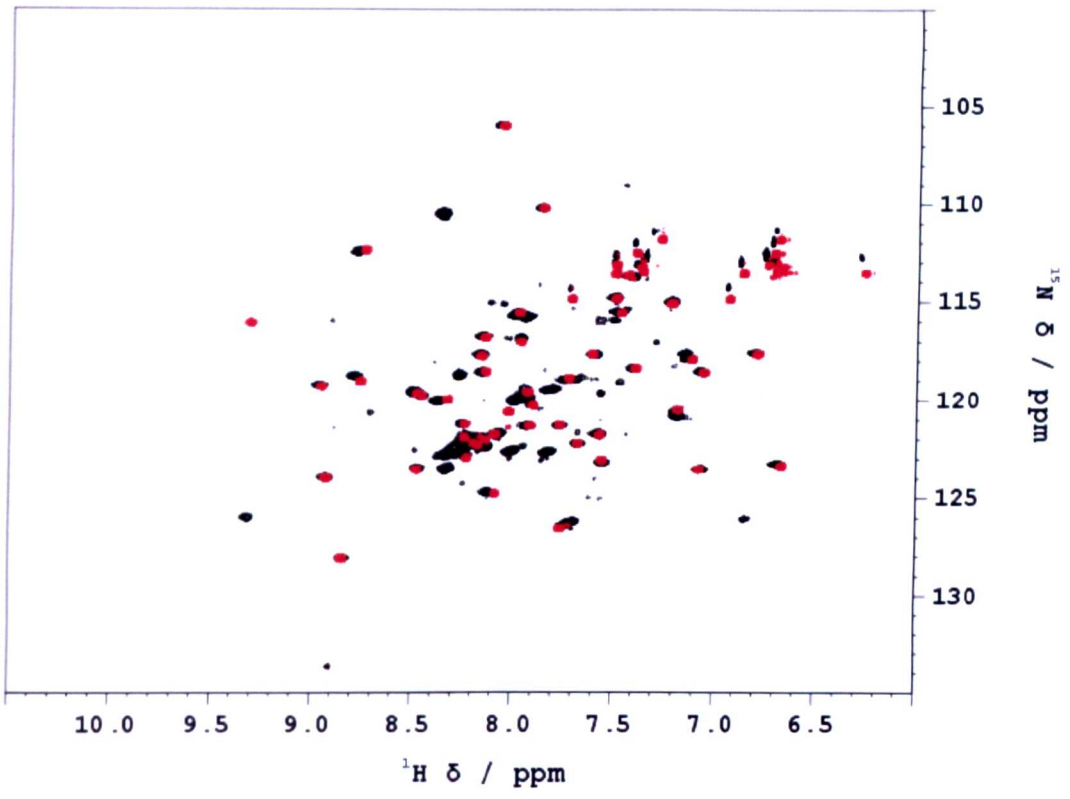


Figure 9.14 - ^{1}H-^{15}N TROSY of the ^{15}N-GST-hHR23A UBA2 (black) and isolated ^{15}N-hHR23A-UBA2 (red). Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

9.2.3.3 Titrations using the GST-hHR23A UBA2

The addition of up to 2 mM unlabelled monoubiquitin and 1 mM linear diubiquitin was added to 500 μM ^{15}N-GST-hHR23A UBA2 (as separate experiments). Concentration dependent chemical shifts consistent with a binding event were observed for the addition of both monoubiquitin and linear diubiquitin. The titration with monoubiquitin showed resonances which were in fast exchange, indicative of a weak binding event (figure 9.16, C). During the titration with monoubiquitin saturation was nearly reached by the addition of 1 mM monoubiquitin (1:2 molar ratio of GST-hHR23A UBA2:monoubiquitin). This suggests a weaker binding interaction than observed for the ^{15}N-GST-p62 UBA with monoubiquitin. The weak interaction is consistent with the weak interaction for isolated hHR23A UBA2 and monoubiquitin (K_d of 400 ± 100 μM)[54]. Both the free and bound forms of the spectrum exhibited resonances with broad signals, consistent with the UBA monomers being conformationally flexible in both forms (figure 9.15, A and B).

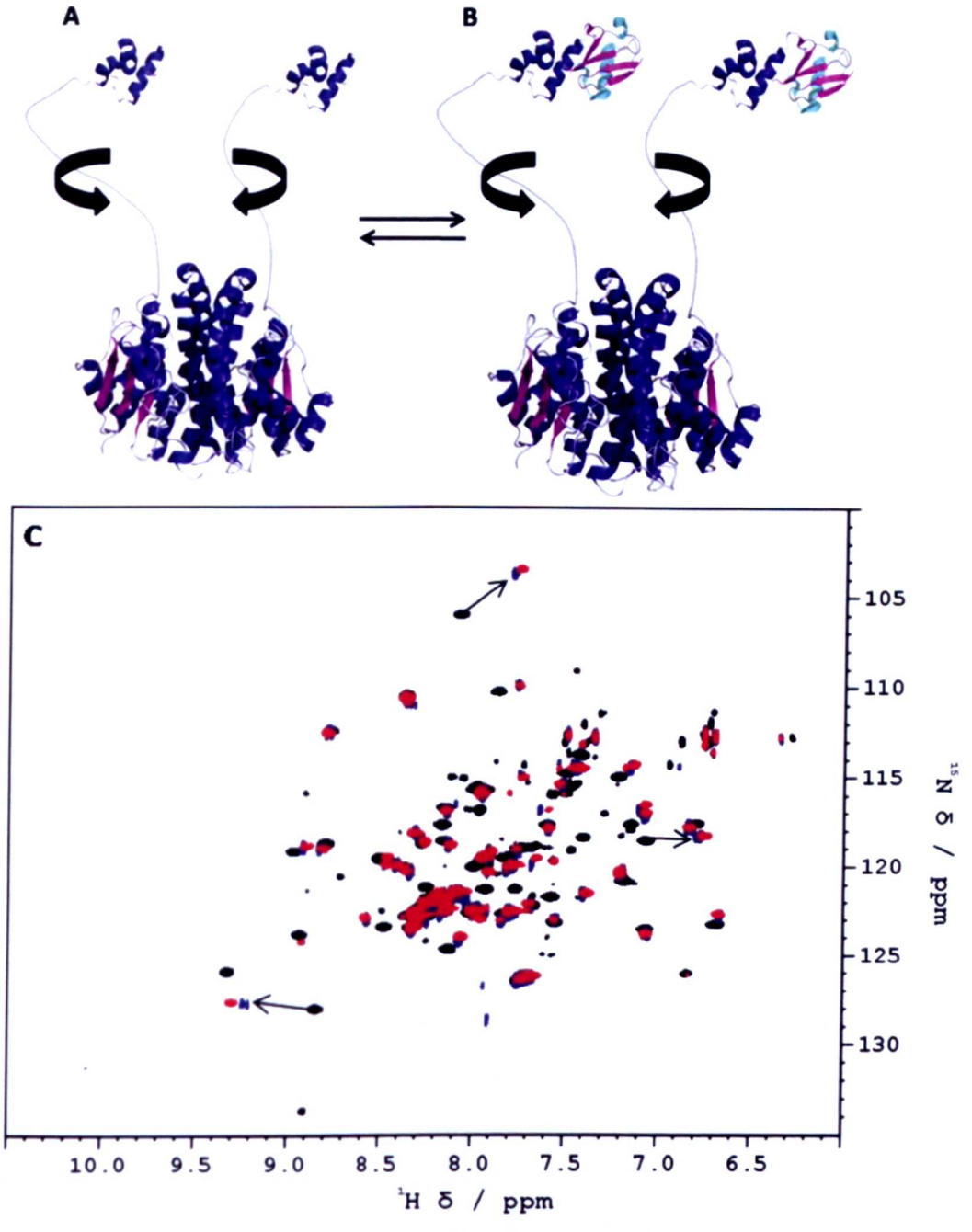


Figure 9.15 - A schematic diagram of the GST-hHR23A UBA2 in A) the free form and B) the bound to monoubiquitin form. The black arrows indicate the flexibility achieved by the linker, which is not restrained by the formation of a UBA dimer at the C-terminus. This flexibility causes broader peaks to appear in the spectrum. C) The ^{1}H-^{15}N TROSY of the ^{15}N-GST-hHR23A UBA2 in the presence of 0 mM (black), 1 mM (blue) and 2 mM (red) monoubiquitin. Arrows are shown for the movement of some of the resonances which had the biggest perturbations between the free and bound forms. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

However when linear diubiquitin was titrated in, resonances which were in intermediate exchange were observed (figure 9.16). After the addition of 50 μM linear diubiquitin significant broadening was seen for resonances in the central region, but these peaks were still able to be detected. Simultaneous to this, the well dispersed peaks on the outer region of the spectrum, which predominantly corresponded to the hHR23A UBA2, had broadened to the point where they were no longer at measurable intensity. The instantaneous broadening of peaks in the titration with linear diubiquitin precluded the analysis of this data, but does strongly suggest an interaction with much higher affinity.

The change from fast to intermediate exchange is indicative of a slower on/off rate between complex formation and dissociation which is associated with a tighter binding interaction. The hHR23A UBA2 domain is well documented to have a preference for Lys48 chains. Although linear chains are not the natural ligand, the data suggests that the GST-fusion has placed the two UBAs in close proximity to favour avid binding to linear diubiquitin. Avid interactions are highly sensitive to the orientation of the binding units as well as the flexibility between the units. The results observed here back up the notion that the linker is not long enough to permit sufficient flexibility between the subunits.

The avidity effects observed by the GST-hHR23A UBA2 are stronger than for the GST-p62 UBA as shown by the change in chemical shift regime. This result was surprising, given that Lys63 and linear chains are favoured by the p62 UBA. On the other hand the result was not totally unexpected given that the GST-hHR23A UBA2 does not have UBA dimerisation competing with ubiquitin binding. Alternatively the length of the polyubiquitin chain could also prove to be influential on avid binding by the p62 UBA, with longer chains required for enhanced avidity effects. The idea of spacer ubiquitins in a polyubiquitin chain to permit the high affinity binding by multiple UBAs is attractive.

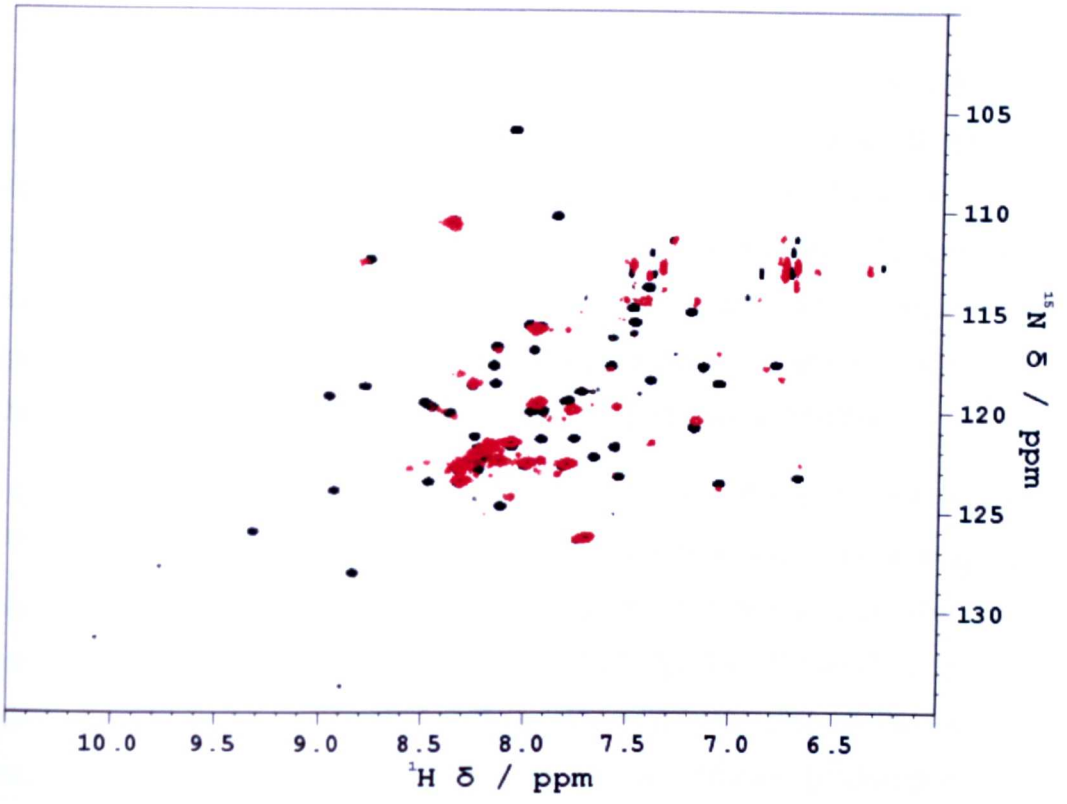


Figure 9.16 - The ^{1}H-^{15}N TROSY of the ^{15}N-GST-hHR23A UBA2 in the free (black) and bound to linear diubiquitin forms (red). Many of the peaks have broadened substantially in the bound form due to the intermediate exchange regime adopted by the interaction. Both spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

9.2.4 GST-p62 T419K dilution

Dilution experiments were conducted in order to see if the UBA in the GST fusion protein was able to dissociate (in the absence of ubiquitin) despite being tethered to the GST dimer. Since low concentrations, which are not ideally suited to NMR, are required to shift the equilibrium towards the monomer in the wild type UBA, a GST-fusion of the weak dimer mutant T419K was used as an alternative. It was postulated that the UBA dimer would not dissociate at lower concentrations because the strength of the GST dimer was so high (1 nm interaction).

A ^{1}H-^{15}N TROSY was recorded at 50 μM intervals from 250 μM to 50 μM for the ^{15}N-GST-p62 T419K. Similar to the cleaved ^{15}N-T419K spectrum, a complicated spectrum with evidence for three species in equilibrium was observed at concentrations greater than 50 μM (figure 9.17, A). The chemical shifts for the three species were near identical to those observed in the cleaved T419K. It was therefore assumed that the ^{15}N-GST-p62 T419K mutation produced the same effects and was able to form the same alternative dimer species. In agreement with the wild type GST-p62 UBA and GST-hHR23A UBA2 spectra, some additional peaks were observed for the ^{15}N-GST-p62 T419K relative to the isolated T419K. Many of these peaks were located at the same chemical shifts as those for the ^{15}N-GST-p62 UBA confirming that they belong to the linker between the GST and the T419K.

During the course of the dilution resonances corresponding to two species (the two UBA dimers) disappeared leaving a spectrum for a single species, the UBA monomer (figure 9.17, B). Spectra recorded at 25 μM and 12.5 μM showed a single species, although the resolution at these concentrations was poor suggesting high protein instability. It was surprising that during the dilution the two dimer species were lost, as it was suspected that tethering to the GST dimer would be able to prevent UBA dissociation. The presence of the GST dimer should in theory maintain the effective local high concentration and thus favour the UBA dimer. Moreover, the short linker should restrict the flexibility of the UBA, again favouring the dimeric state. However, dimer dissociation might be permitted in this case due to the inherently weak dimer formed by the T419K mutant.

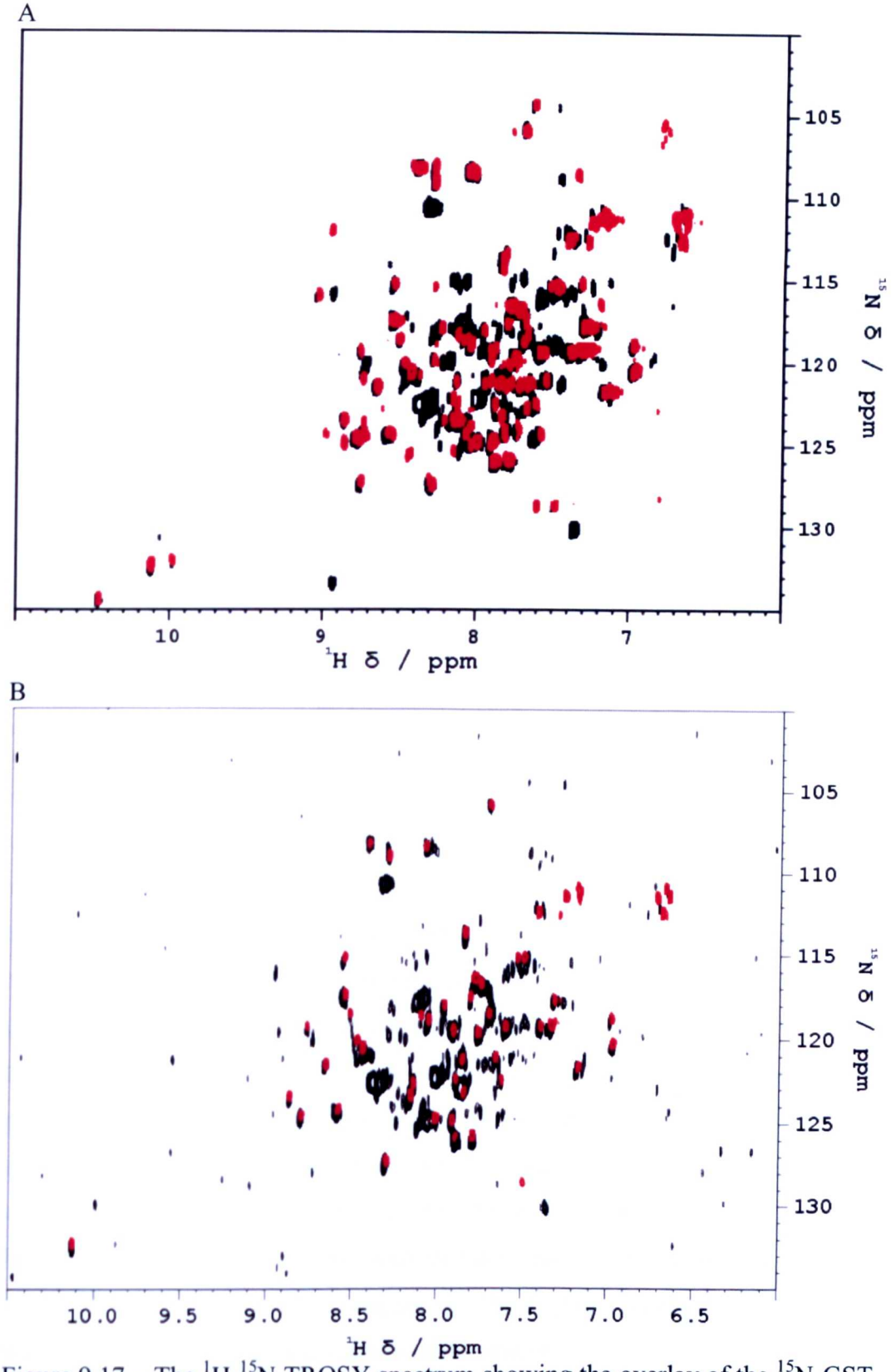


Figure 9.17 – The ^{1}H-^{15}N TROSY spectrum showing the overlay of the ^{15}N-GST-p62 T419K (black) and cleaved T419K (red) at A) 250 μM and B) 50 μM. At 250 μM 3 species are observed, whereas at 50 μM the T419K monomer is observed only. Although the GST-p62 T419K is more unstable at lower concentrations as seen by the poor resolution in the 50 μM spectrum. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 and 298 K.

9.2.5 Monitoring avidity effects by the GST-p62 UBA using nanospray ESI-MS

In order to confirm the findings of the NMR titrations using the GST-p62 UBA, attentions were turned to ESI-MS. Several attempts were made to produce an ESI-MS spectrum; however, signals for the GST-p62 UBA were unable to be detected in the mass spectrometer. Since large proteins and other macromolecular complexes are studied using a nanospray source, this was used as an alternative. Nanospray uses a miniaturized electrospray source and a very low flow rate to generate droplets in the nanometer range. Nanospray has the advantage of more readily preserving noncovalent protein complexes compare to standard ESI-MS methods[288].

Typically higher sample and ammonium acetate concentrations are required for nanospray ESI-MS[176]. Higher ammonium acetate concentrations are able to reduce the effects of non-volatile adduct ions which might be present in the sample, permitting better ionisation of the molecule. Both sample and ammononium acetate concentrations therefore needed to be optimised prior to data collection.

At both 5 and 10 μM the spectrum of the GST-p62 UBA showed predominantly dimer, although a low population of tetramer was also observed. A 1 μM sample was also recorded but a high quality spectrum was unable to be produced that this concentration. The spectrum recorded on a 5 μM sample in 200 μM ammonium acetate was shown to be optimal as it minimised the presence of the tetramer (figure 9.18). Since the presence of the tetramer decreased with concentration it is likely that this is a non specific oligomer which is an artefact of the electrospray process. A mass of 63832.1 Da was observed for the dimer, which was in good agreement with the predicted mass of 63775.6 Da. Spectra at the aforementioned concentrations were recorded in both 100 μM and 200 μM ammonium acetate. A high quality spectrum was observed in both cases, with no obvious changes observed between the two concentrations. The higher concentration of 200 μM was chosen to ensure better ionisation of the protein.

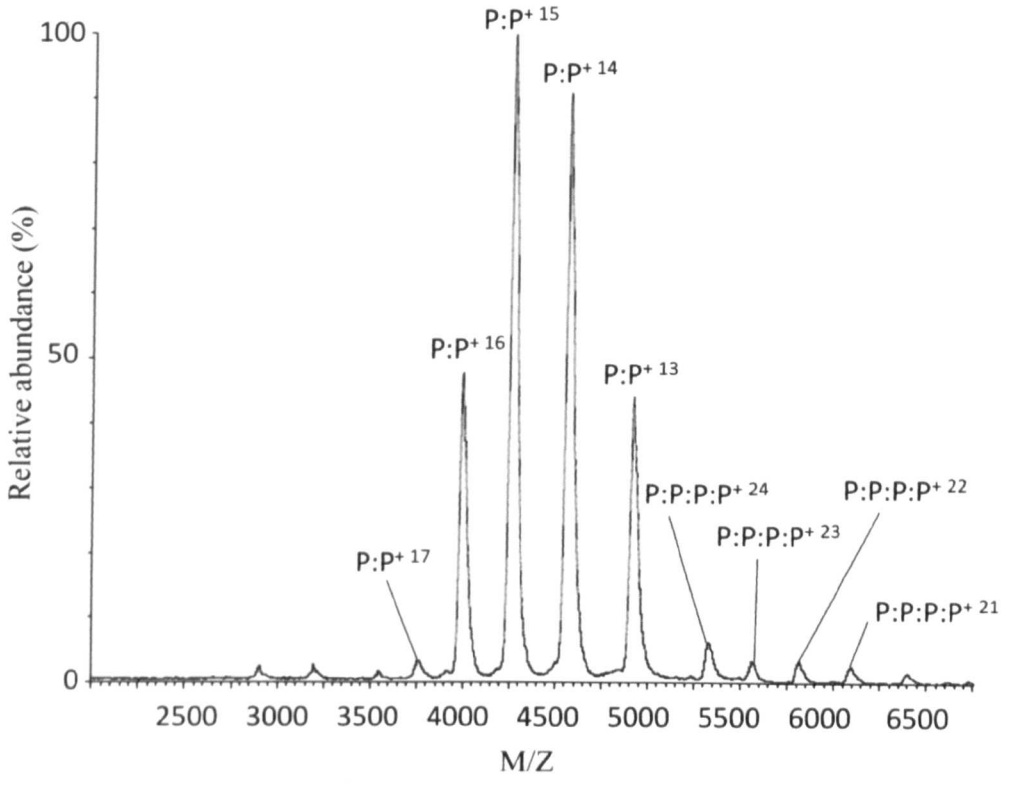


Figure 9.18 - the nanospray ESI-MS spectrum for a 5 μM GST-p62 UBA. A spectrum with predominantly dimer was observed. P:P denotes dimer and P:P:P:P denotes tetramer. The spectrum was recorded in 200 μM ammonium acetate, pH 7.

Once the spectrum for the free GST-p62 UBA had been obtained binding experiments with both monoubiquitin and linear diubiquitin ensued. Experiments with molar ratios of 1:0.5, 1:1 and 1:2 (GST-p62 UBA:ubiquitin) were recorded for both monoubiquitin and linear diubiquitin. In all cases, mass ions were observed for the GST-p62 UBA and the ubiquitin ligand, but not the protein complex (figure 9.19). A range of trap and transfers energies were tried, but a complex was still unable to be detected. Higher voltages are required for bigger complexes but the risk of dissociation is also greater at higher collision energies, especially in weak interactions. Moreover, the backing pressure was also adjusted to see if a higher quality spectrum which contained the complex. All the other ESI-MS binding experiments detailed in this thesis was able to detect very small peaks corresponding to the complex. The poor detection in all cases is linked to the bias of the mass spectrometer against weak hydrophobic interactions. The lack of a complex could be linked to the increased size of the protein as larger protein complexes are more fragile and require greater energy for ionisation.

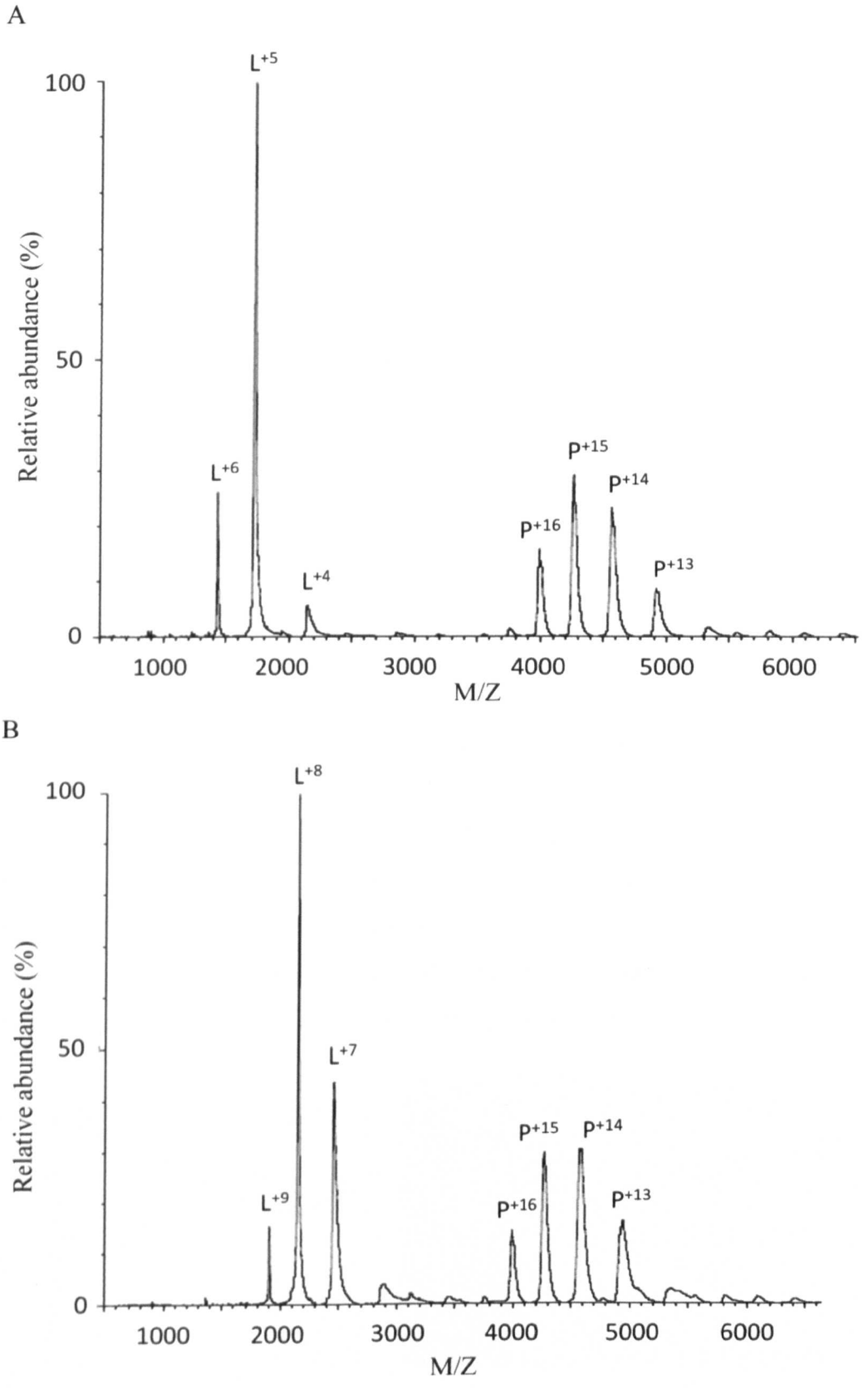


Figure 9.19 – the nanospray ESI-MS spectrum for the GST-p62 UBA when mixed with A) monoubiquitin and B) linear diubiquitin. A molar ratio of 1:1 is shown here. P represents the GST-p62 UBA and L represents either monoubiquitin or linear diubiquitin. Both spectra were recorded in 200 μM ammonium acetate, pH 7.

9.2.6 NMR investigations on the longer GST p62 fusion proteins

9.2.6.1 Degradation is observed in the longer GST p62 fusion proteins

The experiments using the GST p62-UBA construct as a model of the full length p62 protein showed encouraging results. The next step was to perform similar experiments on the longer p62 constructs. The longer constructs would improve the model so that it also contains a long, flexible linker region between the N and C terminal domains. However, the addition of a long linker region N-terminal to the UBA domain has a dramatic effect on the ability to purify a homogenous sample. The longer GST-p62 constructs, GST-p62 261-440, the GST-p62 300-440 and the GST-p62 341-440, all showed significant degradation during the purification protocol when compared to the GST-p62 UBA, even for the GST-p62 341-440 which was the shortest of the longer constructs (figure 9.20). The degradation could be an intrinsic property of the PEST sequence which the linker region encodes. PEST sequences are commonly found in proteins with short half lives[137,138].

The amount of degradation observed by the longer GST fusions prevented accurate measurement of concentration, since the degradation could not be quantified accurately. The presence of the linker in the model produced similar results to attempts to purify the full length protein with mutations in the PB1 domain to prevent self oligomerisation. Since the GST protein is used as a solubility tag it was believed that the GST would have a protective role and help to prevent protein proteolysis; however, this was shown not to be the case.

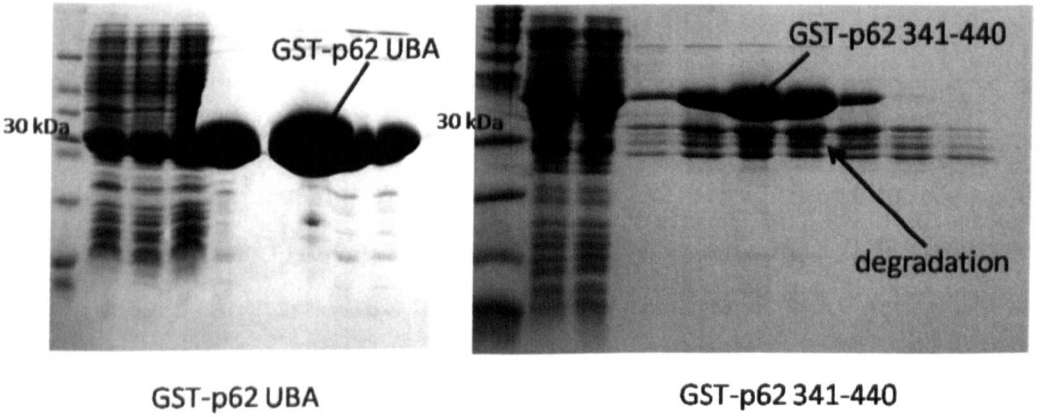


Figure 9.20 – A comparison of the purification gels (both 20 % acrylamide) for the GST-p62 UBA construct and the GST-p62 341-440 construct. The addition of the 41 unstructured residues has a dramatic effect on the amount of degradation exhibited by the GST-p62 341-440 construct.

9.2.6.2 NMR titration studies using the longer p62 GST fusion proteins

The GST-p62 longer constructs were all investigated by NMR despite the observation of significant degradation. As expected, increasing the length of the linker between the GST and the UBA resulted in significantly more complicated spectra. By increasing the unstructured region more sharp peaks become clustered in the central region producing spectra with a high degree of overlap. Moreover, all of the longer GST-p62 constructs all exhibit signs of degradation within their spectra. The spectra produced for all the longer constructs in both the free and bound forms showed resonances corresponding to residues in the UBA which had identical chemical shifts to each other but not to the GST-p62 UBA or isolated p62 UBA. Therefore the GST-p62 261-440 was used as a representative for all the longer constructs. Since ubiquitin binding is the focus of these experiments only residues in the UBA were assigned.

The free ^{1}H-^{15}N TROSY spectrum for the 15N-GST-p62 261-440 showed a wide dispersion of peaks corresponding to the UBA. Most of the resonances for the UBA in the GST-p62 261-440 were present at very similar chemical shifts to those previously reported for the isolated domain (figure 9.21). Although some very slight movement was observed for some resonances. Due to the high correlation between the free forms of the UBA in the isolated domain and the GST-p62 261-440 construct, 41 out of 47 residues in the UBA were able to be assigned based on the previous assignments for the free UBA. The missing residues, namely Ala390, Asp391, Ser399, Leu416, Lys435 and His436, had moved significantly. These residues were surrounded by other peaks which prevented unambiguous assignment. Both Ala390 and As391 are located close to the N-terminus in the UBA construct; however, these residues form part of the flexible linker in the long constructs therefore it was expected that these residues have different chemical shifts as they have changed their chemical environments. Similarly Lys435 and His436 are located at the C-terminus in the UBA construct, but the longer constructs are extended by four residues at the C-terminus. The UBA construct is defined by residues 387-436, but the longer constructs terminate at the natural C-terminus of residue 440. Again the change in chemical environments for these residues can be explained.

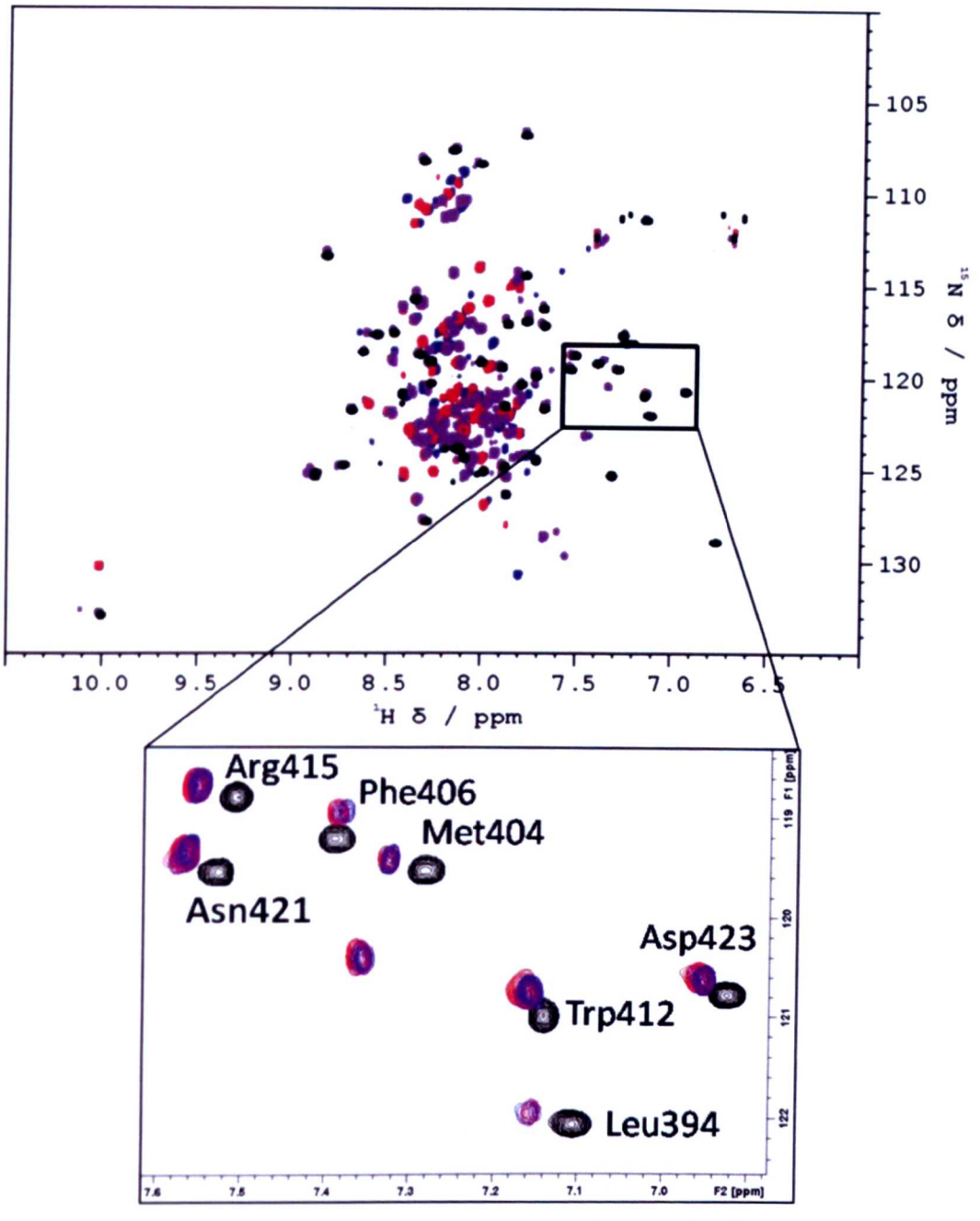


Figure 9.21 – A comparison of the free forms for the longer GST-p62 constructs compared to the isolated ^{15}N-p62 UBA domain. The ^{1}H-^{15}N TROSY for the ^{15}N-GST-p62 261-440 (blue), ^{15}N-GST-p62 300-440 (red), ^{15}N-GST-p62 341-440 (purple) and ^{15}N-p62 UBA. A zoom of the ^{1}H-^{15}N TROSY which highlights resonances which have shifted in the free form for the longer constructs compared to the isolated domain. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

The movement of resonances was much more pronounced in the bound form of the GST-p62 261-440 construct (figure 9.22). Therefore many more peaks were unable to be unambiguously assigned in the bound form compared to the free form. Only 19 residues were able to be assigned based on the chemical shifts previously reported for wild type p62 UBA. The assigned residues were not located in a particular region of the protein, so it is not obvious what has caused these changes. Of the assigned residues mixed signal intensities were observed, suggesting that some of these assignments were not correct. This change in chemical shifts in the longer constructs is consistent with the observation of an alternative bound form when investigating the p62s ability to form a ternary complex (chapter 8).

Titrations with monoubiquitin were completed for all the longer GST-p62 constructs despite being unable to assign the bound form. Up to 1 mM unlabelled ubiquitin was titrated into 250 μM samples of ^{15}N-GST-p62 longer constructs (as separate experiments). Since accurate concentrations could not be determined the titration data could not be examined quantitatively. However, if we assume that the amount of degradation is the same in all the longer constructs and that the extent of degradation remains constant throughout the titration, saturation points can be estimated. Using these assumptions, saturation was observed around 1:2 molar ratio (UBA:ubiquitin). This would imply that the interaction using the longer p62 constructs is stronger, than the isolated UBA domain but weaker than the GST-p62 UBA. This result is consistent with differences in affinity being observed between the isolated UBA and the full length p62 protein using GST pull down assays[170,171] and with the UBA in the GST-p62 UBA construct forming a weaker dimer.

Titrations with linear diubiquitin were not completed due to the poor success shown by the titrations with monoubiqutin. The inability to assign the data combined with the inability to determine accurate protein concentrations meant that limited information could be determined from the data.

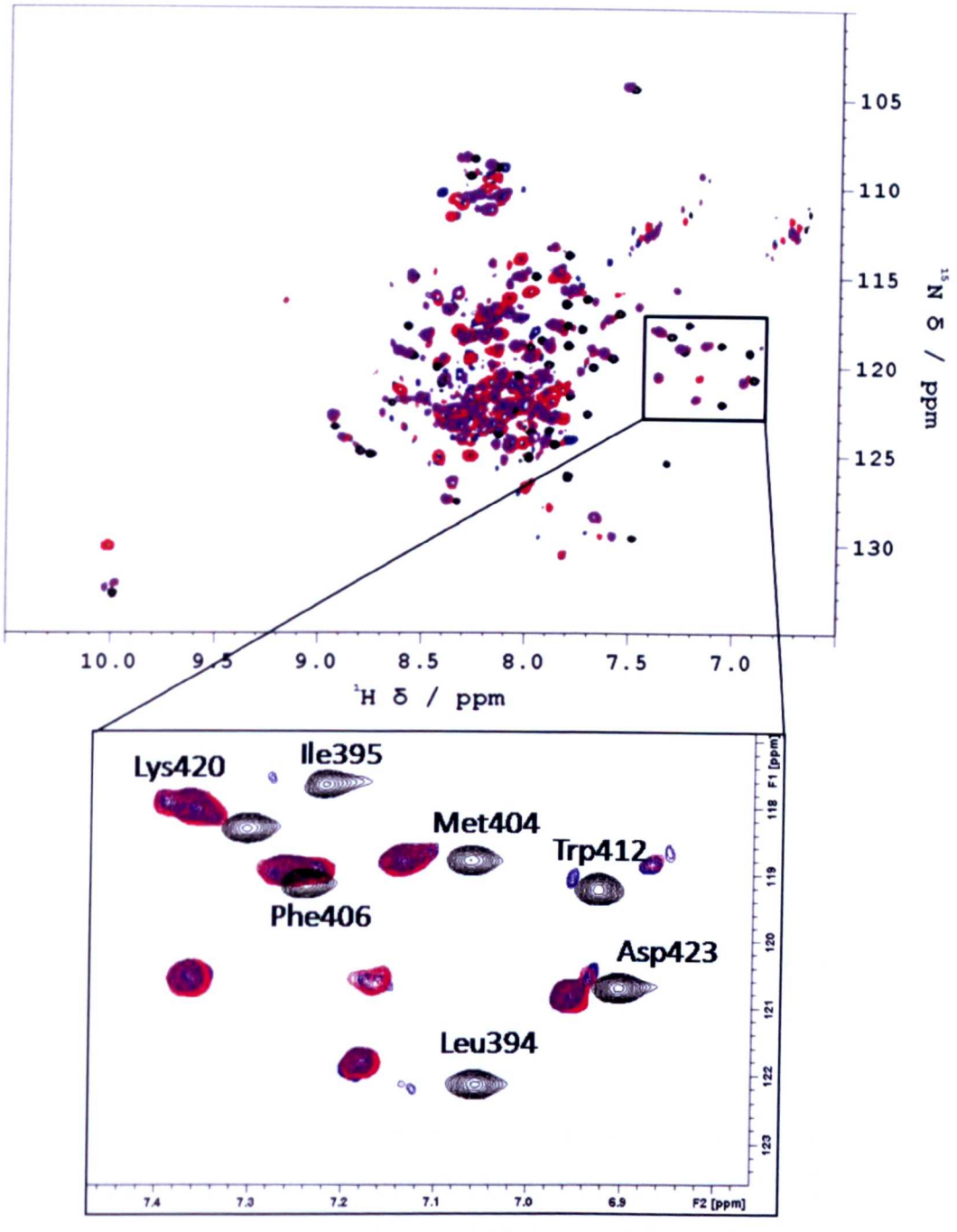


Figure 9.22 – A comparison of the bound forms for the longer GST-p62 constructs compared to the isolated ^{15}N-p62 UBA domain. The ^{1}H-^{15}N TROSY for the ^{15}N-GST-p62 261-440 (blue), ^{15}N-GST-p62 300-440 (red), ^{15}N-GST-p62 341-440 (purple) and ^{15}N-p62 UBA. A zoom of the ^{1}H-^{15}N TROSY which highlights resonances which have shifted in the bound form for the longer constructs compared to the isolated domain. There are no obvious peak matches for Trp412 and Ile395 in the zoom. All spectra were recorded in 25 mM potassium phosphate, 25 mM NaCl, pH 7 at 298 K.

9.3 Discussion

The role of avidity in the binding of polyubiquitin chains to multiple UBDs has only recently begun to be investigated. A study involving 30 UBA domains was able to group UBA domains into four categories according to their linkage preference[104]. It has since been revealed that the results from this study might be incorrect due to the promotion of artificial avidity effects using GST fusion proteins[93]. Since this discovery, the notion that linkage specificity could arise from the multivalent arrangement of UBA domains in oligomeric proteins in nature has begun to be explored.

The p62 protein is capable of self oligomerisation as well as being able to bind to the structurally similar NBR1 protein. The presence of p62 and NBR1 in sequestosomes suggests that multiple UBAs are presented to long polyubiquitin chains. Since the linker is long in both of these proteins enabling conformational flexibility in the UBAs. This permits changes in the orientations of the UBA dependent on the binding partner. Therefore the correct signals are likely to be amplified using avidity effects, suggesting avidity only plays a part in the binding of specific polyubiquitin chains. The specificity of the length of the chain as well as the linkage are likely to be crucial to the high affinity interaction. To date the concept of spacer molecules in polyubiquitin binding by the p62 has not been investigated.

Recently a proposed mechanism of binding by the XIAP protein to linear and Lys63 linked chains was reported using molecular modelling (figure 9.23)[236]. The dimerisation of the RING domain of the XIAP protein facilitates the dimerisation of the XIAP UBA. The XIAP UBA dimer subsequently positions its ubiquitin binding surfaces in an orientation which can simultaneously recognise separate ubiquitin moieties in linear or Lys63 linked polyubiquitin chains to achieve a high affinity interaction. However, the study also revealed that successive ubiquitins were not recognised by the XIAP UBA and that spacer ubiquitin moieties were required. The model was consistent with the weak binding of XIAP to linear linked diubiquitin and the previous report that XIAP requires a tetraubiquitin chain for a detectable interaction.

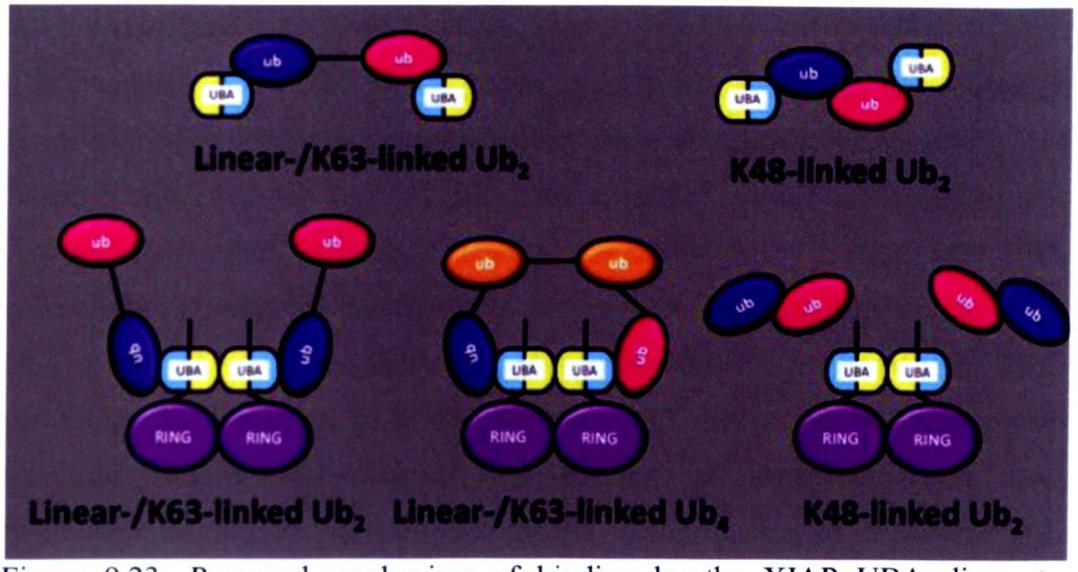


Figure 9.23 -Proposed mechanism of binding by the XIAP UBA dimer to polyubiquitin chains. The ubiquitins which are cloured in orange are the spacer ubiquitins which permit avid interactions by linear or Lys63 linked tetraubiquitin. Taken from Tse *et al*[236].

The XIAP dimer (K_{dim} 900 μM) is weak compared to the p62 UBA dimer (K_{dim} 4.1 μM). In the case of the XIAP, UBA dimerisation is likely to facilitate avid binding interactions. However, the p62 UBA dimer has an inhibitory role in ubiquitin binding due to the partial overlapping of the dimerisation and ubiquitin binding interfaces. To date no other examples of UBA dimers have been shown to have partially overlapping dimerisation and ubiqutin binding interfaces. It is quite clear that the competitive equilibria of the p62 UBA significantly complicates the avidity theory for binding of polyubiquitin chains. Many questions about the role of the UBA dimer in regulating binding affinity remain.

9.4 Conclusions

The attempts to model the full length p62 protein as a dimer and investigate the potential avidity effects in ubiquitin binding were not completely successful. The NMR titration data for the GST-p62 UBA construct showed promising results. Although a dramatic increase in affinity was observed for the GST-p62 UBA compared to the isolated domain. This data suggests that the linker region in the GST-p62 UBA was too short to permit complete flexibility of the UBA, restricting the formation of a high affinity UBA dimer. An interaction with increased affinity was observed for binding to linear diubiquitin over monoubiquitin, suggesting a role for avidity in the binding of oligomeric p62. However, the observed avidity effects were not substantial, with the increase in affinity consistent with there being two binding sites in linear diubiquitin. Although accurate K_d values were unable to be determined from the NMR data at high concentrations, a change in chemical shift regime was not observed in binding to linear diubiquitin, consistent with subtle changes in binding affinity.

The limited role of avidity could be caused by the inhibitory role of UBA dimerisation or the notion that spacer ubiquitin moieties are required for a high affinity interaction. Avidity effects are highly sensitive to both the orientation of the UBA as well as the flexibility of the linker between the subunits. *In vivo* the full length p62 has a longer linker, which could place the UBA domains further apart from each other to favour binding to multiple but not consecutive ubiquitin moieties in a polyubiquitin chain. A greater distance between UBAs is also likely to have an effect on UBA dimerisation by decreasing the effective local concentration thereby decreasing dimer formation.

Complementary reverse titrations whereby GST-p62 UBA was titrated into ^{15}N-monubiquitin and ^{15}N-linear diubiquitin showed confusing results. Saturation was not reached at a molar ratio of 1:2 (ubiquitin:GST-p62 UBA). This result was inconsistent with the forwards titrations which saturated at 1:1 and 1:0.5 molar ratio for monoubiquitin and linear diubiquitin respectively. A higher excess of GST-p62 UBA could not be used as the protein would precitate out of solution. It was not particularly obvious why this occurred as reverse titrations with the isolated UBA and other UBA mutants were able to reach saturation.

The GST-T419K was diluted from 250 μM to 12.5 μM in order to examine if the UBA domain was able to dissociate at lower concentrations in the absence of ubiquitin. The GST fusion of the T419K, like the cleaved T419K, showed the presence of three species at equilibrium. At 50 μM only a single species was observed. This was quite unexpected given that the UBA was tethered to the GST dimer by a short linker. The GST remains dimeric at this concentration and has a K_{dim} of 1 nm. The GST was therefore expected to keep the effective local concentration high and thus favour dimerisation. The data presented for the T419K data, contradicts this theory, although the presence of a weak UBA dimer could affect the results.

Control experiments using a GST fusion of the hHR23A UBA2 protein were also completed. NMR titrations whereby unlabelled moniubiquitin and linear diubiquitin were titrated into ^{15}N GST-hHR23A UBA2, with promising data observed. A weak interaction between the GST-hHR23A UBA2 and monoubiquitin was observed, similar to the weak interaction observed by the isolated hHR23A UBA2 domain and monoubiquitin. The results of this experiment suggest that the increase in affinity observed by the GST-p62 UBA is likely to be linked to the formation of a weaker UBA dimer in the GST-p62 UBA construct, rather than an intrinsic property of the GST fusion protein in the model. The interaction between GST-hHR23A UBA2 and linear diubiquitin showed a change in the chemical shift regime from fast to intermediate exchange. This change is associated with a significant increase in affinity due to slower on and off rates. This data is consistent with previous reports that artificial avidity effects are promoted by the positioning of the two UBAs in close proximity to each other. This data would also suggest that the linker region is this construct is too short to allow total flexibility of the UBA.

However, problems with the model were exposed when a longer linker was inserted. The longer GST-p62 constructs showed significant proteolysis which prevented accurate measurement of protein concentration. The degradation could be linked to the PEST sequences encoded by the linker region. Moreover a combination of chemical shift deviations in the free and bound forms and a highly crowded spectrum prevented the full assignment of the UBA in the bound form. This result was consistent with the observation of an alternative bound form in the cleaved C331S p62 300-440 construct.

The nanospray ESI-MS was unable to detect the complex formed between the GST-p62 UBA and either of the ubiquitin ligands. Despite experimenting with a range of molar ratios of UBA:ubiquitin and varying the parameters, the complex was still not observed. This result was also confusing as nanospray uses more gentle desolvation conditions to preserve noncovalently bond complexes. Although, peaks for the complex are not more abundant for the isolated UBA or any of its mutants, they were able to be detected. It was therefore assumed that the problems associated with detection of the complex were linked to the increased size. In order to get the larger complex to ionise larger collision energies are required but this also increases their fragility as dissociation is more likely to occur.

Overall we were able to show enhanced binding to linear diubiquitin compared to monoubiquitin, consistent with avidity effects increasing the affinity of the interaction. Taken together the data obtained in this chapter showed that in order to generate a successful model a longer linker is required. However, when attempts were made to achieve this degradation became a problem. In this case the degradation could be linked to the fact that PEST signals function as internal degradation signals. We can also conclude from the data that the dimerisation of the p62 UBA is also to oppose the avidity effects. However, we are unable to explore if this inhibitory role is reduced in the full length protein as we were unable to increase the length of the linker.

9.5 Future work

The future of this project lies with generating the NMR structure of the full length p62 protein. The full length p62 protein represents a considerable challenge, considering it contains a large region of unstructured protein. However, recent advances in NMR spectroscopy, such as ^{13}C direct experiments and non uniform sampling, have the potential to solve the structure of the full length protein. The ability to obtain large amounts of a homogenous stable sample appear to be the limiting factor in structural investigation of the full length p62. Since the start of this project the discovery of mutations which enable p62 PB1 monomer and dimers to be formed has been revealed. These mutations were incorporated into the full length protein and attempts to purify it began. However, the purification was unsuccessful for the full length protein when expressed as both a GST fusion or cleaved protein. Degradation was also shown to be a problem in the full length protein. This was combined with a very poor yield, whereby not even 0.5 mg per Litre of growth was obtained in LB media.

Attempts were also made to make another model of the full length p62 protein were also made. This model contained the PB1 domain mutations and removed a large segment of the linker between the ZZ domain and the UBA to minimise degradation. However, the PCR deletion to shorten the linker was unsuccessful despite utilising different conditions and primers. This would not necessarily be the best model as some of the protein binding sites are removed in this linker.

Assuming a suitable model is used or the full length protein is able to be obtained, the binding to longer polyubiquitin chains should also be explored. Or alternatively a molecular model of the full length protein could be used to monitor binding. As in the case for the XIAP UBA, spacer ubiquitin molecules could be required to permit a high affinity interaction for the p62 UBA. The isolated p62 UBA has shown a preference for longer polyubiquitin chains, despite UBA dimerisation opposing the effects of ubiquitin binding. GST pull down assays conducted on the wild type p62 appear to show a preference for tri and tetraubiquitin chains or either Lys63 or Lys48 linkage[140], suggesting that one or two spacer ubiquitins are required to increase the affinity of the interaction. The avidity effects are likely to be amplified in the binding to the preferred length chain as well as linkage.

10 Conclusions

The experiments conducted within this thesis focus on ubiquitin binding by the p62 UBA. Over 30 different mutations in the p62 UBA have been identified in patients with PDB. A loss of ubiquitin binding has been previously shown to be associated with PDB, although the exact mechanisms which underlie the decrease in affinity are currently unknown. A variety of factors including structural alterations and reduced thermal stability have been shown to influence some mutations but not others. Other factors are therefore likely to contribute to the decreased affinity for ubiquitin and subsequently the onset of the disease. We aimed to investigate the role of other factors in ubiquitin binding by the p62 UBA.

The p62 UBA was recently shown to exist as a highly stable dimer, which is unusual for a UBA domain. In fact the p62 has been shown to have a novel mechanism of ubiquitin binding, whereby dimerisation competes with ubiquitin binding as the two interfaces partially overlap. The p62 UBA dimer dissociates to facilitate the interaction of the monomer with ubiquitin. The interaction between the p62 UBA monomer and ubiquitin is weak (40 ± 10 μM). The role of the dimer in regulating the affinity of ubiquitin was explored using mutations along the dimerisation interface (chapter 6). A phosphorylation site at Ser403 in the UBA was also recently identified. The role of phosphorylation in regulating ubiquitin binding was explored structurally and biophysically using phosphomimetic mutations (chapter 7). Subtle changes in affinity were observed for the mutants compared to the wild type p62 UBA by NMR (table 10.1 and 10.2)

In chapter 6 we introduced mutations to solvent exposed residues in helix 2 which forms the dimerisation interface. Weaker dimers were formed by the T414A, T414K and T419K mutants. K_{dim} values of 141 ± 1 μM, 146 ± 1.5 μM and 165 ± 30 μM were calculated from ITC dilution studies for the T414A, T414K and T419K mutants respectively. The wild type p62 UBA had a K_{dim} of 4.1 ± 0.6 μM, meaning that the T414A, T414K and T419K mutants were an order of magnitude weaker than the wild type p62 UBA. We were able to show that there is a relationship between weaker dimers and the affinity of the interaction with ubiquitin. The weak dimer mutants were shown to have an increased affinity for ubiquitin.

The T419K mutant had shifted the equilibrium furthest as the monomeric species was visible at high concentration (500 μM). This mutation enabled the study of the monomeric species at low concentration (<50 μM). ^{15}N heteronuclear NOE experiments revealed an average value of 0.62 indicative of a structure which exhibits some flexibility. This mutant also reached full saturation before the T414A and T414K at 1:1 molar ratio. However, when a K_d for binding by the T419K monomer was calculated the affinity was found to be weaker than the value previously calculated for wild type. This is likely to be linked to an inaccurate value for K_{dim} calculated by a poorly fitted dissociation curve.

In chapter 7 we generated the phosphomimetic mutants S403D and S403E and the non-phosphorylated mutant S403A. We were able to show a preference for the phosphomimetic mutants compared to the S403A control. K_d values of 71.4 ± 5.5 μM, 45.5 ± 2.1 μM and 27.8 ± 0.8 μM were calculated for the S403A, S403D and S403E mutants using ESI-MS. Although these values are likely to be affected by the bias of the mass spectrometer against hydrophobic interactions. A competitive binding experiment using ESI-MS also revealed a preference for the phosphomimetic mutants over the S403A mutant. Reverse NMR titrations whereby the S403 mutants were titrated into ^{15}N-ubiquitin revealed a larger binding interface on ubiquitin for the phosphomimetic mutants. K_d's for the interaction were calculated at 10.3 ± 0.5 μM, 7.1 ± 0.2 μM and 8.7 ± 0.4 μM for the S403A, S403D and S403E mutants respectively. Consistent with the K_d's calculated by ESI-MS, a preference for the phosphomimetic mutants was observed. A forward NMR titration for the S403D mutant produced a K_d of 39.1 ± 4.7 μM, which was highly similar to the value reported for wild type p62 UBA.

We were also able to confirm the monomer-dimer equilibrium was not affected by the S403 mutation. We were able to show that the structural integrity of the dimer was similar to that of the wild type and therefore changes in affinity could be attributed to the role of phosphorylation only. K_{dim} values of 3.5 ± 0.5 μM, 6.7 ± 0.6 μM and 4.2 ± 0.5 μM were calculated for the S403A, S403D and S403E mutants respectively. Overall the impact of phosphorylation alone on the UBA was minimal, as shown by the calculated K_d values, it was suspected that the small changes in affinity could be amplified by avidity effects when binding to oligomeric p62.

Table 10.1 – K_d values for p62 UBA as calculated from forward NMR titrations.

P62 UBA monomer	K_{dim}	K_{obs}	K_d
Wild type	4.1 ± 0.6	741 ± 80	40 ± 10
T419K	165 ± 30	185.9 ± 8.7	175.1 ± 20.4
S403D	6.7 ± 0.6	229 ± 34	39.1 ± 4.7

Table 10.2 – K_d values for p62 UBA ascalculated from reverse NMR titrations.

	K_{dim}	K_{obs}	K_d
Wild type	4.1 ± 0.6	540 ± 45	47.1 ± 10
S403A	3.5 ± 0.5	373 ± 19	10.3 ± 0.5
S403D	6.7 ± 0.6	337 ± 11	7.1 ± 0.2
S403E	4.2 ± 0.5	319 ± 11	8.7 ± 0.4

The role of ubiquitin binding in the context of the full length protein was also investigated in this thesis. Differences in affinity have been observed between the isolated UBA domain and the full length protein, suggesting that factors outside of the UBA are also involved in regulating the binding of the p62 UBA. The structure of the full length p62 protein has yet to be determined and is considered to be a long term goal in understanding the role of the p62 as a scaffold protein which is able to regulate various signalling pathways simultaneously. Chapters 8 and 9 investigated ubiquitin binding in longer constructs of the p62.

In chapter 8 the p62s ability to bind to multiple proteins was explored using a construct of the p62 which extended back from the C-terminus to residue 300. This construct encoded both the LIR and UBA of the p62. The p62 has been shown to target ubiquitinated proteins for degradation by either the UPS or autophagy. We were able to form a p62 mediated ternary complex with MAP-LC3 and ubiquitin using NMR, ESI-MS and molecular modelling techniques. An interaction between the MAP-LC3 and ubiquitin once bound to p62 was not observed, despite a flexible linker separating the LIR and UBA. A higher affinity interaction was observed for binding to MAP-LC3 over ubiquitin, consistent with the previous literature. The binding of both proteins occurred irrelevant of occupation of the other binding site. We were also unable to observe an allosteric relationship between MAP-LC3 and ubiquitin in p62 binding. However, the possibility that allostery could have a role in the binding of longer polyubiquitin chains can not be excluded, as longer chains could block access to the LIR. One

particularly unusual observation was that a different ubiquitin bound form was observed in the C331S p62 300-440 construct as chemical shift deviations were observed for residues in the UBA.

The flexibility of the linker was probed using ^{1}H-^{15}N heteronuclear NOE experiments. With only MAP-LC3 bound to the C331S p62 300-440 a large number of negative peaks were observed suggesting the high flexibility in the linker was maintained despite protein binding. Although when ubiquitin and both proteins were bound nearly all residues produced positive signals highlighting a less flexible structure. It would therefore seem that ubiquitin binding has a greater effect than MAP-LC3 on the flexibility of the p62. The molecular model revelled that the linker possessed a large number of pronounced kinks due a high prevalence of Pro residues. These Pro residues also restrict movement of the linker and are therefore likely to prevent an interaction between the two bound proteins.

In chapter 9 the role of avidity in raising the affinity of the interaction to a high affinity interaction were explored using a model of the full length p62 protein. The model consisted of fragments of the p62 extending back from the C-terminus expressed as GST fusion proteins. The p62 protein has been shown to be highly oligomeric due to its ability to self olgiomerise via the N-terminal PB1 domain. *In vivo* the p62 is found in speckles which cause effective local high concentrations. The p62 is therefore capable of presenting an array of UBAs to long polyubiquitin chains. Since the interaction of a single UBA to ubiquitin has been shown to be weak (40 ± 10 μM), it was postulated that the affinity of the interaction could be levered to a physiologically relevant affinity using avidity effects. NMR spectroscopy was used as predominantly signals corresponding to the UBA were visible, with the GST component being largely 'NMR silent'.

The GST-p62 UBA construct bound to ubiquitin much quicker than expected, suggesting that a weaker dimer was formed UBA. This could have been caused by the short linker restricting the ability to forma high affinity dimer. Avidity effects were observed in the binding of the GST-p62 UBA to linear diubiquitin compared to monoubiquitin. Control experiments using the monomeric hHR23A UBA2 domain were also completed. A higher affinity interaction was observed by the

GST-hHR23A UBA2 in binding to monoubiquitin, compared to the isolated domain, suggesting again that the linker was too short to permit total conformational flexibility. Much stronger avidity effects were observed in the binding of GST- hHR23A UBA2 to linear diubiquitin than for GST-p62 UBA. This is consistent with dimerisation not competing with binding and that the short linker has placed the two UBA monomers close in space to favour avid binding to linear diubiquitin. It was deduced that the linker was too short in these constructs. However, when experiments using GST fusions of p62 constructs with longer linker regions were completed some additional problems with the model were observed. The predicted protective effect of the GST to prevent degradation in the linker was not observed. In addition, similar to the cleaved longer C331S p62 300-440 construct, a different ubiquitin bound form was observed.

Overall it is clear that the differences in affinity are exerted by the full length protein relative to the isolated UBA domain. In order to fully understand ubiquitin recognition and binding by the p62, the future should lie with the full length protein. Thermal stability, small structural perturbations, dimerisation and phosphorylation have all been shown to have a role in regulating the affinity of the UBA, although it is clear that other factors which are currently unknown also have a role in ubiquitin binding. Moreover, how these factors are able to act in concert to regulate the affinity. It is also unknown if the UBA domain is dimeric in the full length protein. *In vivo* the effective local concentration of UBA in oligomeric p62 can be decreased given that a long linker separates the ZZ domain and the UBA. Although we have explored ubiquitin binding in the isolated UBA and in longer p62 constructs, many questions remain about ubiquitin recognition and regulation of binding by the p62 UBA.

11 References

1. Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M., BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Research* 2010, 38, D750-D753.
2. Bhattacharyya, R. P.; Remenyi, A.; Yeh, B. J.; Lim, W. A., Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual Review of Biochemistry* 2006, 75, 655-680.
3. Scott, J. D.; Pawson, T., Cell Signaling in Space and Time: Where Proteins Come Together and When They're Apart. *Science* 2009, 326, 1220-1224.
4. Good, M. C.; Zalatan, J. G.; Lim, W. A., Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* 2011, 332, 680-686.
5. Komander, D., The emerging complexity of protein ubiquitination. *Biochemical Society Transactions* 2009, 37, 937-953.
6. Giles, J., Chemistry Nobel for trio who revealed molecular death-tag. *Nature* 2004, 431, 729-729.
7. Vijaykumar, S.; Bugg, C. E.; Wilkinson, K. D.; Cook, W. J., 3-dimensional structure of ubiquitin at 2.8 a resolution. *Proceedings of the National Academy of Sciences of the United States of America* 1985, 82, 3582-3585.
8. Vijaykumar, S.; Bugg, C. E.; Cook, W. J., structure of ubiquitin refined at 1.8 a resolution. *Journal of Molecular Biology* 1987, 194, 531-544.
9. Perica, T.; Chothia, C., Ubiquitin - molecular mechanisms for recognition of different structures. *Current Opinion in Structural Biology* 2010, 20, 367-376.
10. Vijaykumar, S.; Bugg, C. E.; Wilkinson, K. D.; Vierstra, R. D.; Hatfield, P. M.; Cook, W. J., comparison of the 3-dimensional structures of human, yeast, and oat ubiquitin. *Journal of Biological Chemistry* 1987, 262, 6396-6399.
11. Pickart, C. M.; Eddins, M. J., Ubiquitin: structures, functions, mechanisms. *Biochimica Et Biophysica Acta-Molecular Cell Research* 2004, 1695, 55-72.
12. Hershko, A.; Ciechanover, A., The ubiquitin system. *Annual Review of Biochemistry* 1998, 67, 425-479.
13. Lange, O. F.; Lakomek, N.-A.; Fares, C.; Schroeder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmueller, H.; Griesinger, C.; de Groot, B. L., Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 2008, 320, 1471-1475.
14. Sloper-Mould, K. E.; Jemc, J. C.; Pickart, C. M.; Hicke, L., Distinct functional surface regions on ubiquitin. *Journal of Biological Chemistry* 2001, 276, 30483-30489.
15. Bomar, M. G.; D'Souza, S.; Bienko, M.; Dikic, I.; Walker, G. C.; Zhou, P., Unconventional Ubiquitin Recognition by the Ubiquitin-Binding Motif within the Y Family DNA Polymerases iota and Rev1. *Molecular Cell* 2010, 37, 408-417.
16. Garner, T. P.; Strachan, J.; Shedden, E. C.; Long, J. E.; Cavey, J. R.; Shaw, B.; Layfield, R.; Searle, M. S., Independent interactions of ubiquitin-binding domains in a ubiquitin-mediated ternary complex. *Biochemistry* 2011, 50, 9076-87.
17. Wlodarski, T.; Zagrovic, B., Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106, 19346-19351.
18. Pelzer, C.; Kassner, I.; Matentzoglu, K.; Singh, R. K.; Wollscheid, H. P.; Scheffner, M.; Schmidtke, G.; Groettrup, M., UBE1L2, a novel E1 enzyme specific for ubiquitin. *Journal of Biological Chemistry* 2007, 282, 23010-23014.
19. Pickett, J., Post-translational modification - UBE1, you're not alone. *Nature Reviews Molecular Cell Biology* 2007, 8, 599-599.

20. Schulman, B. A.; Harper, J. W., Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nature Reviews Molecular Cell Biology* 2009, 10, 319-331.
21. Markson, G.; Kiel, C.; Hyde, R.; Brown, S.; Charalabous, P.; Bremm, A.; Semple, J.; Woodsmith, J.; Duley, S.; Salehi-Ashtiani, K.; Vidal, M.; Komander, D.; Serrano, L.; Lehner, P.; Sanderson, C. M., Analysis of the human E2 ubiquitin conjugating enzyme protein interaction network. *Genome Research* 2009, 19, 1905-1911.
22. David, Y.; Ziv, T.; Admon, A.; Navon, A., The E2 Ubiquitin-conjugating Enzymes Direct Polyubiquitination to Preferred Lysines. *Journal of Biological Chemistry* 2010, 285, 8595-8604.
23. Deshaies, R. J.; Joazeiro, C. A. P., RING Domain E3 Ubiquitin Ligases. *Annual Review of Biochemistry* 2009, 78, 399-434.
24. Borden, K. L. B., RING domains: Master builders of molecular scaffolds? *Journal of Molecular Biology* 2000, 295, 1103-1112.
25. Koegl, M.; Hoppe, T.; Schlenker, S.; Ulrich, H. D.; Mayer, T. U.; Jentsch, S., A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* 1999, 96, 635-644.
26. Hoppe, T., Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. *Trends in Biochemical Sciences* 2005, 30, 183-187.
27. Richly, H.; Rape, M.; Braun, S.; Rumpf, S.; Hoege, C.; Jentsch, S., A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* 2005, 120, 73-84.
28. Haenzelmann, P.; Stingele, J.; Hofmann, K.; Schindelin, H.; Raasi, S., The Yeast E4 Ubiquitin Ligase Ufd2 Interacts with the Ubiquitin-like Domains of Rad23 and Dsk2 via a Novel and Distinct Ubiquitin-like Binding Domain. *Journal of Biological Chemistry* 2010, 285, 20390-20398.
29. Cook, W. J.; Jeffrey, L. C.; Carson, M.; Chen, Z. J.; Pickart, C. M., structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *Journal of Biological Chemistry* 1992, 267, 16467-16471.
30. Hicke, L., Protein regulation by monoubiquitin. *Nature Reviews Molecular Cell Biology* 2001, 2, 195-201.
31. Di Fiore, P. P.; Polo, S.; Hofmann, K., When ubiquitin meets ubiquitin receptors: a signalling connection. *Nature Reviews Molecular Cell Biology* 2003, 4, 491-497.
32. Lai, Z. H.; Ferry, K. V.; Diamond, M. A.; Wee, K. E.; Kim, Y. B.; Ma, J. H.; Yang, T.; Benfield, P. A.; Copeland, R. A.; Auger, K. R., Human mdm2 mediates multiple mono-ubiquitination of p53 by a mechanism requiring enzyme isomerization. *Journal of Biological Chemistry* 2001, 276, 31357-31367.
33. Matsumoto, M. L.; Wickliffe, K. E.; Dong, K. C.; Yu, C.; Bosanac, I.; Bustos, D.; Phu, L.; Kirkpatrick, D. S.; Hymowitz, S. G.; Rape, M.; Kelley, R. F.; Dixit, V. M., K11-Linked Polyubiquitination in Cell Cycle Control Revealed by a K11 Linkage-Specific Antibody. *Molecular Cell* 2010, 39, 477-484.
34. Bremm, A.; Freund, S. M. V.; Komander, D., Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature Structural & Molecular Biology* 2010, 17, 939-U47.
35. Komander, D.; Reyes-Turcu, F.; Licchesi, J. D. F.; Odenwaelder, P.; Wilkinson, K. D.; Barford, D., Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *Embo Reports* 2009, 10, 466-473.
36. Rohaim, A.; Kawasaki, M.; Kato, R.; Dikic, I.; Wakatsuki, S., Structure of a compact conformation of linear diubiquitin. *Acta Crystallographica Section D-Biological Crystallography* 2012, 68, 102-108.

37. Kumar, K. S. A.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A., Total Chemical Synthesis of Di-ubiquitin Chains. *Angewandte Chemie-International Edition* 2010, 49, 9126-9131.
38. Spasser, L.; Kumar, K. S. A.; Erlich, L. A.; Bavikar, S. N.; Brik, A., Total Chemical Synthesis of All di-Ubiquitin Chains. *Biopolymers* 2011, 96, 525-525.
39. Weikart, N. D.; Sommer, S.; Mootz, H. D., Click synthesis of ubiquitin dimer analogs to interrogate linkage-specific UBA domain binding. *Chemical Communications* 2012, 48, 296-298.
40. Xu, P.; Duong, D. M.; Seyfried, N. T.; Cheng, D.; Xie, Y.; Robert, J.; Rush, J.; Hochstrasser, M.; Finley, D.; Peng, J., Quantitative Proteomics Reveals the Function of Unconventional Ubiquitin Chains in Proteasomal Degradation. *Cell* 2009, 137, 133-145.
41. Thrower, J. S.; Hoffman, L.; Rechsteiner, M.; Pickart, C. M., Recognition of the polyubiquitin proteolytic signal. *Embo Journal* 2000, 19, 94-102.
42. Chen, Z. J.; Sun, L. J., Nonproteolytic Functions of Ubiquitin in Cell Signaling. *Molecular Cell* 2009, 33, 275-286.
43. Chen, Z. J.; Pickart, C. M., a 25-kilodalton ubiquitin carrier protein (e2) catalyzes multi-ubiquitin chain synthesis via lysine-48 of ubiquitin. *Journal of Biological Chemistry* 1990, 265, 21835-21842.
44. Haldeman, M. T.; Xia, G.; Kasperek, E. M.; Pickart, C. M., Structure and function of ubiquitin conjugating enzyme E2-25K: The tail is a core-dependent activity element. *Biochemistry* 1997, 36, 10526-10537.
45. Cook, W. J.; Jeffrey, L. C.; Kasperek, E.; Pickart, C. M., structure of tetraubiquitin shows how multiubiquitin chains can be formed. *Journal of Molecular Biology* 1994, 236, 601-609.
46. Eddins, M. J.; Varadan, R.; Flushman, D.; Pickart, C. M.; Wolberger, C., Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. *Journal of Molecular Biology* 2007, 367, 204-211.
47. Trempe, J. F.; Brown, N. R.; Noble, M. E. M.; Endicott, J. A., A new crystal form of Lys48-linked diubiquitin. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* 2010, 66, 994-998.
48. Phillips, C. L.; Thrower, J.; Pickart, C. M.; Hill, C. P., Structure of a new crystal form of tetraubiquitin. *Acta Crystallographica Section D-Biological Crystallography* 2001, 57, 341-344.
49. Lai, M. Y.; Zhang, D.; Laronde-Leblanc, N.; Fushman, D., Structural and biochemical studies of the open state of Lys48-linked diubiquitin. *Biochim Biophys Acta* 2012.
50. Ryabov, Y.; Fushman, D., Interdomain mobility in Di-ubiquitin revealed by NMR. *Proteins-Structure Function and Bioinformatics* 2006, 63, 787-796.
51. Ryabov, Y. E.; Fushman, D., A model of interdomain mobility in a multidomain protein. *Journal of the American Chemical Society* 2007, 129, 3315-3327.
52. Beal, R.; Deveraux, Q.; Xia, G.; Rechsteiner, M.; Pickart, C., Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proceedings of the National Academy of Sciences of the United States of America* 1996, 93, 861-866.
53. Varadan, R.; Walker, O.; Pickart, C.; Fushman, D., Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology* 2002, 324, 637-647.
54. Varadan, R.; Assfalg, M.; Raasi, S.; Pickart, C.; Fushman, D., Structural determinants for selective recognition of a lys48-linked polyubiquitin chain by a UBA domain. *Molecular Cell* 2005, 18, 687-698.
55. Hofmann, R. M.; Pickart, C. M., In vitro assembly and recognition of Lys-63 polyubiquitin chains. *Journal of Biological Chemistry* 2001, 276, 27936-27943.

56. VanDemark, A. P.; Hofmann, R. M.; Tsui, C.; Pickart, C. M.; Wolberger, C., Molecular insights into polyubiquitin chain assembly: Crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 2001, 105, 711-720.
57. Varadan, R.; Assfalg, M.; Haririnia, A.; Raasi, S.; Pickart, C.; Fushman, D., Solution conformation of Lys(63)-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *Journal of Biological Chemistry* 2004, 279, 7055-7063.
58. Weeks, S. D.; Grasty, K. C.; Hernandez-Cuebas, L.; Loll, P. J., Crystal structures of Lys-63-linked tri- and di-ubiquitin reveal a highly extended chain architecture. *Proteins-Structure Function and Bioinformatics* 2009, 77, 753-759.
59. Datta, A. B.; Hura, G. L.; Wolberger, C., The Structure and Conformation of Lys63-Linked Tetraubiquitin. *Journal of Molecular Biology* 2009, 392, 1117-1124.
60. Tenno, T.; Fujiwara, K.; Tochio, H.; Iwai, K.; Morita, E. H.; Hayashi, H.; Murata, S.; Hiroaki, H.; Sato, M.; Tanaka, K.; Shirakawa, M., Structural basis for distinct roles of Lys63- and Lys48-linked polyubiquitin chains. *Genes to Cells* 2004, 9, 865-875.
61. Chiu, Y. H.; Zhao, M.; Chen, Z. J., Ubiquitin in NF-kappa B Signaling. *Chemical Reviews* 2009, 109, 1549-1560.
62. Sato, Y.; Yoshikawa, A.; Yamagata, A.; Mimura, H.; Yamashita, M.; Ookata, K.; Nureki, O.; Iwai, K.; Komada, M.; Fukai, S., Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains. *Nature* 2008, 455, 358-U19.
63. Newton, K.; Matsumoto, M. L.; Wertz, I. E.; Kirkpatrick, D. S.; Lill, J. R.; Tan, J.; Dugger, D.; Gordon, N.; Sidhu, S. S.; Fellouse, F. A.; Komuves, L.; French, D. M.; Ferrando, R. E.; Lam, C.; Compaan, D.; Yu, C.; Bosanac, I.; Hymowitz, S. G.; Kelley, R. F.; Dixit, V. M., Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* 2008, 134, 668-678.
64. Dammer, E. B.; Na, C. H.; Xu, P.; Seyfried, N. T.; Duong, D. M.; Cheng, D. M.; Gearing, M.; Rees, H.; Lah, J. J.; Levey, A. I.; Rush, J.; Peng, J. M., Polyubiquitin Linkage Profiles in Three Models of Proteolytic Stress Suggest the Etiology of Alzheimer Disease. *Journal of Biological Chemistry* 2011, 286, 10457-10465.
65. Wu, T.; Merbl, Y.; Huo, Y.; Gallop, J. L.; Tzur, A.; Kirschner, M. W., UBE2S drives elongation of K11-linked ubiquitin chains by the Anaphase-Promoting Complex. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107, 1355-1360.
66. Garnett, M. J.; Mansfeld, J.; Godwin, C.; Matsusaka, T.; Wu, J.; Russell, P.; Pines, J.; Venkitaraman, A. R., UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. *Nature Cell Biology* 2009, 11, 1363-U241.
67. Williamson, A.; Wickliffe, K. E.; Mellone, B. G.; Song, L.; Karpen, G. H.; Rape, M., Identification of a physiological E2 module for the human anaphase-promoting complex. *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106, 18213-18218.
68. Castaneda, C. A.; Kashyap, T.; Fushman, D., K11-linked Diubiquitin Exhibits Significant Interdomain Dynamics. *Biophysical Journal* 2011, 100, 375-375.
69. Bremm, A.; Komander, D., Emerging roles for Lys11-linked polyubiquitin in cellular regulation. *Trends in Biochemical Sciences* 2011, 36, 355-363.
70. Rape, M.; Reddy, S. K.; Kirschner, M. W., The processivity of multiubiquitination by the APC determines the order of substrate degradation. *Cell* 2006, 124, 89-103.
71. Jin, L.; Williamson, A.; Banerjee, S.; Philipp, I.; Rape, M., Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* 2008, 133, 653-665.
72. Song, L.; Rape, M., Regulated Degradation of Spindle Assembly Factors by the Anaphase-Promoting Complex. *Molecular Cell* 2010, 38, 369-382.
73. Kirisako, T.; Kamei, K.; Murata, S.; Kato, M.; Fukumoto, H.; Kanie, M.; Sano, S.; Tokunaga, F.; Tanaka, K.; Iwai, K., A ubiquitin ligase complex assembles linear polyubiquitin chains. *Embo Journal* 2006, 25, 4877-4887.

74. Rahighi, S.; Ikeda, F.; Kawasaki, M.; Akutsu, M.; Suzuki, N.; Kato, R.; Kensche, T.; Uejima, T.; Bloor, S.; Komander, D.; Randow, F.; Wakatsuki, S.; Dikic, I., Specific Recognition of Linear Ubiquitin Chains by NEMO Is Important for NF-kappa B Activation. *Cell* 2009, 136, 1098-1109.
75. Kensche, T.; Tokunaga, F.; Ikeda, F.; Goto, E.; Iwai, K.; Dikic, I., Analysis of Nuclear Factor-kappa B (NF-kappa B) Essential Modulator (NEMO) Binding to Linear and Lysine-linked Ubiquitin Chains and Its Role in the Activation of NF-kappa B. *Journal of Biological Chemistry* 2012, 287, 23626-23634.
76. Iwai, K.; Tokunaga, F., Linear polyubiquitination: a new regulator of NF-kappa B activation. *Embo Reports* 2009, 10, 706-713.
77. Tokunaga, F.; Sakata, S.-i.; Saeki, Y.; Satomi, Y.; Kirisako, T.; Kamei, K.; Nakagawa, T.; Kato, M.; Murata, S.; Yamaoka, S.; Yamamoto, M.; Akira, S.; Takao, T.; Tanaka, K.; Iwai, K., Involvement of linear polyubiquitylation of NEMO in NF-kappa B activation. *Nature Cell Biology* 2009, 11, 123-U40.
78. Virdee, S.; Ye, Y.; Nguyen, D. P.; Komander, D.; Chin, J. W., Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nature Chemical Biology* 2010, 6, 750-757.
79. Hurley, J. H.; Lee, S.; Prag, G., Ubiquitin-binding domains. *Biochemical Journal* 2006, 399, 361-372.
80. Dikic, I.; Wakatsuki, S.; Walters, K. J., Ubiquitin-binding domains - from structures to functions. *Nature Reviews Molecular Cell Biology* 2009, 10, 659-671.
81. Winget, J. M.; Mayor, T., The Diversity of Ubiquitin Recognition: Hot Spots and Varied Specificity. *Molecular Cell* 2010, 38, 627-635.
82. Bomar, M. G.; Pai, M. T.; Tzeng, S. R.; Li, S. S. C.; Zhou, P., Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta. *Embo Reports* 2007, 8, 247-251.
83. Lee, S.; Tsai, Y. C.; Mattera, R.; Smith, W. J.; Kostelansky, M. S.; Weissman, A. M.; Bonifacino, J. S.; Hurley, J. H., Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. *Nature Structural & Molecular Biology* 2006, 13, 264-271.
84. Kulathu, Y.; Akutsu, M.; Bremm, A.; Hofmann, K.; Komander, D., Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *Nature Structural & Molecular Biology* 2009, 16, 1328-1330.
85. Hirano, S.; Kawasaki, M.; Ura, H.; Kato, R.; Raiborg, C.; Stenmark, H.; Wakatsuki, S., Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nature Structural & Molecular Biology* 2006, 13, 272-277.
86. Bosanac, I.; Wertz, I. E.; Pan, B.; Yu, C.; Kusam, S.; Lam, C.; Phu, L.; Phung, Q.; Maurer, B.; Arnott, D.; Kirkpatrick, D. S.; Dixit, V. M.; Hymowitz, S. G., Ubiquitin Binding to A20 ZnF4 Is Required for Modulation of NF-kappa B Signaling. *Molecular Cell* 2010, 40, 548-557.
87. Lo Conte, L.; Chothia, C.; Janin, J., The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* 1999, 285, 2177-2198.
88. Ye, Y.; Blaser, G.; Horrocks, M. H.; Ruedas-Rama, M. J.; Ibrahim, S.; Zhukov, A. A.; Orte, A.; Klenerman, D.; Jackson, S. E.; Komander, D., Ubiquitin chain conformation regulates recognition and activity of interacting proteins. *Nature* 2012, 492, 266-270.
89. Fushman, D.; Wilkinson, K. D., Structure and recognition of polyubiquitin chains of different lengths and linkage. *F1000 biology reports* 2011, 3, 26-26.
90. Long, D.; Brueschweiler, R., In Silico Elucidation of the Recognition Dynamics of Ubiquitin. *Plos Computational Biology* 2011, 7.
91. Reyes-Turcu, F. E.; Wilkinson, K. D., Polyubiquitin Binding and Disassembly By Deubiquitinating Enzymes. *Chemical Reviews* 2009, 109, 1495-1508.

92. Markin, C. J.; Wei, X.; Spyracopoulos, L., *Mechanism for Recognition of Polyubiquitin Chains: Balancing Affinity through Interplay between Multivalent Binding and Dynamics. Journal of the American Chemical Society* 2010, 132, 11247-11258.
93. Sims, J. J.; Haririnia, A.; Dickinson, B. C.; Fushman, D.; Cohen, R. E., Avid interactions underlie the Lys63-linked polyubiquitin binding specificities observed for UBA domains. *Nature Structural & Molecular Biology 2009, 16, 883-U112.*
94. Komander, D.; Clague, M. J.; Urbe, S., *Breaking the chains: structure and function of the deubiquitinases. Nature Reviews Molecular Cell Biology* 2009, 10, 550-563.
95. Dieckmann, T.; Withers-Ward, E. S.; Jarosinski, M. A.; Liu, C. F.; Chen, I. S. Y.; Feigon, J., Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nature Structural Biology* 1998, 5, 1042-1047.
96. Withers-Ward, E. S.; Mueller, T. D.; Chen, I. S. Y.; Feigon, J., Biochemical and structural analysis of the interaction between the UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr. *Biochemistry* 2000, 39, 14103-14112.
97. Yuan, X. M.; Simpson, P.; McKeown, C.; Kondo, H.; Uchiyama, K.; Wallis, R.; Dreveny, I.; Keetch, C.; Zhang, X. D.; Robinson, C.; Freemont, P.; Matthews, S., Structure, dynamics and interactions of p47, a major adaptor of the AAA ATPase, p97. *Embo Journal* 2004, 23, 1463-1473.
98. Swanson, K. A.; Hicke, L.; Radhakrishnan, I., Structural basis for monoubiquitin recognition by the Ede1 UBA domain. *Journal of Molecular Biology* 2006, 358, 713-724.
99. Zhang, D.; Raasi, S.; Fushman, D., Affinity makes the difference: Nonselective interaction of the UBA domain of ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains. *Journal of Molecular Biology* 2008, 377, 162-180.
100. Ohno, A.; Jee, J.; Fujiwara, K.; Tenno, T.; Goda, N.; Tochio, H.; Kobayashi, H.; Hiroaki, H.; Shirakawa, M., Structure of the UBA domain of Dsk2p in complex with ubiquitin: Molecular determinants for ubiquitin recognition. *Structure* 2005, 13, 521-532.
101. Trempe, J. F.; Brown, N. R.; Lowe, E. D.; Gordon, C.; Campbell, I. D.; Noble, M. E. M.; Endicott, J. A., Mechanism of Lys48-linked polyubiquitin chain recognition by the Mud1 UBA domain. *Embo Journal* 2005, 24, 3178-3189.
102. Nguyen, L.; Kozlov, G.; Lin, T.; De Crescenzo, G.; Park, M.; Gehring, K., Structural basis of ubiquitin recognition by the UBA domain of the EDD ubiquitin ligase. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* 2008, 86, 188-188.
103. Wilkinson, C. R. M.; Seeger, M.; Hartmann-Petersen, R.; Stone, M.; Wallace, M.; Semple, C.; Gordon, C., Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nature Cell Biology* 2001, 3, 939-943.
104. Raasi, S.; Varadan, R.; Fushman, D.; Pickart, C. M., Diverse polyubiquitin interaction properties of ubiquitin-associated domains. *Nature Structural & Molecular Biology* 2005, 12, 708-714.
105. Nijman, S. M. B.; Luna-Vargas, M. P. A.; Velds, A.; Brummelkamp, T. R.; Dirac, A. M. G.; Sixma, T. K.; Bernards, R., A genomic and functional inventory of deubiquitinating enzymes. *Cell* 2005, 123, 773-786.
106. Reyes-Turcu, F. E.; Ventii, K. H.; Wilkinson, K. D., Regulation and Cellular Roles of Ubiquitin-Specific Deubiquitinating Enzymes. *Annual Review of Biochemistry* 2009, 78, 363-397.
107. Komander, D.; Barford, D., Structure of the A20 OTU domain and mechanistic insights into deubiquitination. *Biochemical Journal* 2008, 409, 77-85.
108. Drag, M.; Mikolajczyk, J.; Bekes, M.; Reyes-Turcu, F. E.; Ellman, J. A.; Wilkinson, K. D.; Salvesen, G. S., Positional-scanning fluorigenic substrate libraries reveal unexpected specificity determinants of DUBs (deubiquitinating enzymes). *Biochemical Journal* 2008, 415, 367-375.

109. Hu, M.; Li, P. W.; Song, L.; Jeffrey, P. D.; Chernova, T. A.; Wilkinson, K. D.; Cohen, R. E.; Shi, Y. G., Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *Embo Journal* 2005, 24, 3747-3756.
110. Komander, D.; Lord, C. J.; Scheel, H.; Swift, S.; Hofmann, K.; Ashworth, A.; Barford, D., The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Molecular Cell* 2008, 29, 451-464.
111. Reyes-Turcu, F. E.; Horton, J. R.; Mullally, J. E.; Heroux, A.; Cheng, X. D.; Wilkinson, K. D., The ubiquitin binding domain ZnFUBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* 2006, 124, 1197-1208.
112. Puls, A.; Schmidt, S.; Grawe, F.; Stabel, S., Interaction of protein kinase C zeta with ZIP, a novel protein kinase c-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* 1997, 94, 6191-6196.
113. Duran, A.; Serrano, M.; Leitges, M.; Flores, J. M.; Picard, S.; Brown, J. P.; Moscat, J.; Diaz-Meco, M. T., The atypical PKC-interacting protein p62 is an important mediator of RANK-activated osteoclastogenesis. *Developmental Cell* 2004, 6, 303-309.
114. Moscat, J.; Diaz-Meco, M. T., p62 at the Crossroads of Autophagy, Apoptosis, and Cancer. *Cell* 2009, 137, 1001-1004.
115. Babu, J. R.; Geetha, T.; Wooten, M. W., Sequestosome 1/p62 shuttles polyubiquitinated tau for proteasomal degradation. *Journal of Neurochemistry* 2005, 94, 192-203.
116. Pankiv, S.; Clausen, T. H.; Lamark, T.; Brech, A.; Bruun, J.-A.; Outzen, H.; Overvatn, A.; Bjorkoy, G.; Johansen, T., p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *Journal of Biological Chemistry* 2007, 282, 24131-24145.
117. Ichimura, Y.; Kumanomidou, T.; Sou, Y. S.; Mizushima, T.; Ezaki, J.; Ueno, T.; Kominami, E.; Yamane, T.; Tanaka, K.; Komatsu, M., Structural basis for sorting mechanism of p62 in selective autophagy. *Journal of Biological Chemistry* 2008, 283, 22847-22857.
118. Layfield, R.; Hocking, L. J., SQSTM1 and Paget's disease of bone. *Calcified Tissue International* 2004, 75, 347-357.
119. Layfield, R.; Searle, M. S., Disruption of ubiquitin-mediated processes in diseases of the brain and bone. *Biochemical Society Transactions* 2008, 36, 469-471.
120. Sanz, L.; Diaz-Meco, M. T.; Nakano, H.; Moscat, J., The atypical PKC-interacting protein p62 channels NF-kappa B activation by the IL-1-TRAF6 pathway. *Embo Journal* 2000, 19, 1576-1586.
121. Wooten, M. W.; Seibenhener, M. L.; Mamidipudi, V.; Diaz-Meco, M. T.; Barker, P. A.; Moscat, J., The atypical protein kinase C-interacting protein p62 is a scaffold for NF-kappa B activation by nerve growth factor. *Journal of Biological Chemistry* 2001, 276, 7709-7712.
122. Zatloukal, K.; Stumptner, C.; Fuchsbichler, A.; Heid, H.; Schnoelzer, M.; Kenner, L.; Kleinert, R.; Prinz, M.; Aguzzi, A.; Denk, H., p62 is a common component of cytoplasmic inclusions in protein aggregation diseases. *American Journal of Pathology* 2002, 160, 255-263.
123. Gal, J.; Strom, A. L.; Kilty, R.; Zhang, F. J.; Zhu, H. N., p62 accumulates and enhances aggregate formation in model systems of familial amyotrophic lateral sclerosis. *Journal of Biological Chemistry* 2007, 282, 11068-11077.
124. Moscat, J.; Diaz-Meco, M. T.; Albert, A.; Campuzano, S., Cell signaling and function organized by PB1 domain interactions. *Molecular Cell* 2006, 23, 631-640.
125. Saio, T.; Yokochi, M.; Inagaki, F., The NMR structure of the p62 PB1 domain, a key protein in autophagy and NF-kappa B signaling pathway. *Journal of Biomolecular NMR* 2009, 45, 335-341.

126. Saio, T.; Yokochi, M.; Kumeta, H.; Inagaki, F., PCS-based structure determination of protein-protein complexes. *Journal of Biomolecular Nmr* 2010, 46, 271-280.
127. Muller, S.; Kursula, I.; Zou, P.; Wilmanns, M., Crystal structure of the PB1 domain of NBR1. *Febs Letters* 2006, 580, 341-344.
128. Seibenhener, M. L.; Babu, J. R.; Geetha, T.; Wong, H. C.; Krishna, N. R.; Wooten, M. W., Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. *Molecular and Cellular Biology* 2004, 24, 8055-8068.
129. Kang, Y.; Vossler, R. A.; Diaz-Martinez, L. A.; Winter, N. S.; Clarke, D. J.; Walters, K. J., UBL/UBA ubiquitin receptor proteins bind a common tetraubiquitin chain. *Journal of Molecular Biology* 2006, 356, 1027-1035.
130. Kang, Y.; Zhang, N. X.; Koepp, D. M.; Walters, K. J., Ubiquitin receptor proteins hHR23a and hPLIC2 interact. *Journal of Molecular Biology* 2007, 365, 1093-1101.
131. Isogai, S.; Morimoto, D.; Arita, K.; Unzai, S.; Tenno, T.; Hasegawa, J.; Sou, Y.; Komatsu, M.; Tanaka, K.; Shirakawa, M.; Tochio, H., Crystal Structure of the Ubiquitin-associated (UBA) Domain of p62 and Its Interaction with Ubiquitin. *Journal of Biological Chemistry* 2011, 286, 31864-31874.
132. Ponting, C. P.; Blake, D. J.; Davies, K. E.; KendrickJones, J.; Winder, S. J., ZZ and TAZ: New putative zinc fingers in dystrophin and other proteins. *Trends in Biochemical Sciences* 1996, 21, 11-13.
133. Sanz, L.; Sanchez, P.; Lallena, M. J.; Diaz-Meco, M. T.; Moscat, J., The interaction of p62 with RIP links the atypical PKCs to NF-kappa B activation. *Embo Journal* 1999, 18, 3044-3053.
134. Zheng, Y. T.; Shahnazari, S.; Brech, A.; Lamark, T.; Johansen, T.; Brumell, J. H., The Adaptor Protein p62/SQSTM1 Targets Invading Bacteria to the Autophagy Pathway. *Journal of Immunology* 2009, 183, 5909-5916.
135. Wooten, M. W.; Geetha, T.; Seibehener, M. L.; Babu, J. R.; Diaz-Meco, M. T.; Moscat, J., The p62 scaffold regulates nerve growth factor-induced NF-kappa B activation by influencing TRAF6 polyubiquitination. *Journal of Biological Chemistry* 2005, 280, 35625-35629.
136. Lamothe, B.; Webster, W. K.; Gopinathan, A.; Besse, A.; Campos, A. D.; Darnay, B. G., TRAF6 ubiquitin ligase is essential for RANKL signaling and osteoclast differentiation. *Biochemical and Biophysical Research Communications* 2007, 359, 1044-1049.
137. Rogers, S.; Wells, R.; Rechsteiner, M., amino-acid-sequences common to rapidly degraded proteins - the pest hypothesis. *Science* 1986, 234, 364-368.
138. Rechsteiner, M.; Rogers, S. W., PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences* 1996, 21, 267-271.
139. Ciani, B.; Layfield, R.; Cavey, J. R.; Sheppard, P. W.; Searle, M. S., Structure of the ubiquitin-associated domain of p62 (SQSTM1) and implications for mutations that cause Paget's disease of bone. *Journal of Biological Chemistry* 2003, 278, 37409-37412.
140. Long, J.; Gallagher, T. R. A.; Cavey, J. R.; Sheppard, P. W.; Ralston, S. H.; Layfield, R.; Searle, M. S., Ubiquitin recognition by the ubiquitin-associated domain of p62 involves a novel conformational switch. *Journal of Biological Chemistry* 2008, 283, 5427-5440.
141. Long, J.; Garner, T. P.; Pandya, M. J.; Craven, C. J.; Chen, P.; Shaw, B.; Williamson, M. P.; Layfield, R.; Searle, M. S., Dimerisation of the UBA Domain of p62 Inhibits Ubiquitin Binding and Regulates NF-kappa B Signalling. *Journal of Molecular Biology* 2010, 396, 178-194.
142. Seibenhener, M. L.; Geetha, T.; Wooten, M. W., Sequestosome 1/p62 - More than just a scaffold. *Febs Letters* 2007, 581, 175-179.
143. Heinen, C.; Garner, T. P.; Long, J.; Bottcher, C.; Ralston, S. H.; Cavey, J. R.; Searle, M. S.; Layfield, R.; Dantuma, N. P., Mutant p62/SQSTM1 UBA domains linked to Paget's

disease of bone differ in their abilities to function as stabilization signals. *Febs Letters* 2010, 584, 1585-1590.
144. Heinen, C.; Acs, K.; Hoogstraten, D.; Dantuma, N. P., C-terminal UBA domains protect ubiquitin receptors by preventing initiation of protein degradation. *Nature Communications* 2011, 2.
145. Karin, M.; Yamamoto, Y.; Wang, Q. M., The IKKNF-kappa B system: A treasure trove for drug development. *Nature Reviews Drug Discovery* 2004, 3, 17-26.
146. Xu, J.; Wu, H. F.; Ang, E. S. M.; Yip, K.; Woloszyn, M.; Zheng, M. H.; Tan, R. X., NF-kappa B modulators in osteolytic bone diseases. *Cytokine & Growth Factor Reviews* 2009, 20, 7-17.
147. Chan, F. K.-M., Three is better than one: Pre-ligand receptor assembly in the regulation of TNF receptor signaling. *Cytokine* 2007, 37, 101-107.
148. Chen, Z. J.; Sun, L.; Deng, L.; Kanayama, A.; Seth, R.; Xia, Z.; Ea, C., Ubiquitin signaling in the NF-kappaB pathway. *Molecular Biology of the Cell* 2004, 15, 119A-119A.
149. Wang, C.; Deng, L.; Hong, M.; Akkaraju, G. R.; Inoue, J.; Chen, Z. J. J., TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* 2001, 412, 346-351.
150. Mizukami, J.; Takaesu, G.; Akatsuka, H.; Sakurai, H.; Ninomiya-Tsuji, J.; Matsumoto, K.; Sakurai, N., Receptor activator of NF-kappa B ligand (RANKL) activates TAK1 mitogen-activated protein kinase kinase kinase through a signaling complex containing RANK, TAB2, and TRAF6. *Molecular and Cellular Biology* 2002, 22, 992-1000.
151. Crockett, J. C.; Mellis, D. J.; Scott, D. I.; Helfrich, M. H., New knowledge on critical osteoclast formation and activation pathways from study of rare genetic diseases of osteoclasts: focus on the RANK/RANKL axis. *Osteoporosis International* 2011, 22, 1-20.
152. Phan, T. C. A.; Xu, J.; Zheng, M. H., Interaction between osteoblast and osteoclast: impact in bone disease. *Histology and Histopathology* 2004, 19, 1325-1344.
153. Cooper, C.; Schafheutle, K.; Dennison, E.; Kellingray, S.; Guyer, P.; Barker, D., The epidemiology of Paget's disease in Britain: Is the prevalence decreasing? *Journal of Bone and Mineral Research* 1999, 14, 192-197.
154. Hansen, M. F.; Seton, M.; Merchant, A., Osteosarcoma in Paget's disease of bone. *Journal of Bone and Mineral Research* 2006, 21, P58-P63.
155. Chung, P. Y. J.; Beyens, G.; Guanabens, N.; Boonen, S.; Papapoulos, S.; Karperien, M.; Eekhoff, M.; Van Wesenbeeck, L.; Jennes, K.; Geusens, P.; Offeciers, E.; Van Offel, J.; Westhovens, R.; Zmierczak, H.; Devogelaer, J. P.; Van Hul, W., Founder effect in different European countries for the recurrent P392L SQSTM1 mutation in Paget's disease of bone. *Calcified Tissue International* 2008, 83, 34-42.
156. Rea, S. L.; Walsh, J. P.; Ward, L.; Magno, A. L.; Ward, B. K.; Shaw, B.; Layfield, R.; Kent, G. N.; Xu, J.; Ratajczak, T., Sequestosome 1 mutations in Paget's disease of bone in Australia: Prevalence, genotype/phenotype correlation and a novel non-UBA domain mutation (P364S) associated with increased NF-kappa B signalling without loss of ubiquitin-binding. *Bone* 2009, 44, 088.
157. Cody, J. D.; Singer, F. R.; Roodman, G. D.; Otterund, B.; Lewis, T. B.; Leppert, M.; Leach, R. J., Genetic linkage of Paget disease of the bone to chromosome 18q. *American Journal of Human Genetics* 1997, 61, 1117-1122.
158. Ralston, S. H.; Langston, A. L.; Reid, I. R., Pathogenesis and management of Paget's disease of bone. *Lancet* 2008, 372, 155-163.
159. Laurin, N.; Brown, J. P.; Lemainque, A.; Duchesne, A.; Huot, D.; Lacourciere, Y.; Drapeau, G.; Verreault, J.; Raymond, V.; Morissette, J., Paget disease of bone: Mapping of two loci at 5q35-qter and 5q31. *American Journal of Human Genetics* 2001, 69, 528-543.
160. Hocking, L. J.; Herbert, C. A.; Nicholls, R. K.; Williams, F.; Bennett, S. T.; Cundy, T.; Nicholson, G. C.; Wuyts, W.; Van Hul, W.; Ralston, S. H., Genomewide search in familial Paget disease of bone shows evidence of genetic heterogeneity with candidate loci on

chromosomes 2q36, 10p13, and 5q35. *American Journal of Human Genetics* 2001, 69, 1055-1061.
161. Laurin, N.; Brown, J. P.; Morissette, J.; Raymond, V., Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget disease of bone. *American Journal of Human Genetics* 2002, 70, 1582-1588.
162. Johnson-Pais, T. L.; Wisdom, J. H.; Weldon, K. S.; Cody, J. D.; Hansen, M. F.; Singer, F. R.; Leach, R. J., Three novel mutations in SQSTM1 identified in familial Paget's disease of bone. *Journal of Bone and Mineral Research* 2003, 18, 1748-1753.
163. Falchetti, A.; Di Stefano, M.; Marini, F.; Del Monte, F.; Mavilia, C.; Strigoli, D.; De Feo, M. L.; Isaia, G.; Masi, L.; Amedei, A.; Cioppi, F.; Ghinoi, V.; Bongi, S. M.; Di Fede, G.; Sferrazza, C.; Rini, G. B.; Melchiorre, D.; Matucci-Cerinic, M.; Brandi, M. L., Two novel mutations at exon 8 of the sequestosome 1 (SQSTM1) gene in an Italian series of patients affected by Paget's disease of bone (PDB). *Journal of Bone and Mineral Research* 2004, 19, 1013-1017.
164. Eekhoff, E. W. M.; Karperien, M.; Houtsma, D.; Zwinderman, A. H.; Dragoiescu, C.; Kneppers, A. L. J.; Papapoulos, S. E., Familial Paget's disease in The Netherlands - Occurrence, identification of new mutations in the sequestosome 1 gene, and their clinical associations. *Arthritis and Rheumatism* 2004, 50, 1650-1654.
165. Hocking, L. J.; Lucas, G. J. A.; Daroszewska, A.; Cundy, T.; Nicholson, G. C.; Donath, J.; Walsh, J. P.; Finlayson, C.; Cavey, J. R.; Ciani, B.; Sheppard, P. W.; Searle, M. S.; Layfield, R.; Ralston, S. H., Novel UBA domain mutations of SQSTM1 in Paget's disease of bone: Genotype phenotype correlation, functional analysis, and structural consequences. *Journal of Bone and Mineral Research* 2004, 19, 1122-1127.
166. Falchetti, A.; Di Stefano, M.; Marini, F.; Ortolani, S.; Ulivieri, M.; Bergui, S.; Masi, L.; Cepollaro, C.; Benucci, M.; Di Munno, O.; Rossini, M.; Adami, S.; Del Puente, A.; Isaia, G.; Torricelli, F.; Brandi, M.; GenePage, P., Genetic Epidemiology of Paget's Disease of Bone in Italy: sequestosome1/p62 Gene Mutational Test and Haplotype Analysis at 5q35 in a Large Representative Series of Sporadic and Familial Italian Cases of Paget's Disease of Bone. *Calcified Tissue International* 2009, 84, 20-37.
167. Gennari, L.; Gianfrancesco, F.; Di Stefano, M.; Rendina, D.; Merlotti, D.; Esposito, T.; De Paola, V.; Aloia, A.; Martini, G.; Mazzetti, M.; Gallone, S.; Raincro, I.; Pinessi, L.; Isaia, G.; Strazzullo, P.; Nuti, R.; Mossetti, G., Large Collaborative Study on Geographic Variation of SQSTM1 Mutations in Paget's Disease of Bone in Italy. *Journal of Bone and Mineral Research* 2008, 23, S223-S223.
168. Visconti, M. R.; Langston, A. L.; Alonso, N.; Goodman, K.; Selby, P. L.; Fraser, W. D.; Ralston, S. H., Mutations of SQSTM1 Are Associated With Severity and Clinical Outcome in Paget Disease of Bone. *Journal of Bone and Mineral Research* 2010, 25, 2368-2373.
169. Rea, S. L.; Walsh, J. P.; Ward, L.; Yip, K.; Ward, B. K.; Kent, G. N.; Steer, J. H.; Xu, J. K.; Ratajczak, T., A novel mutation (K378X) in the sequestosome 1 gene associated with increased NF-kappa B signaling and Paget's disease of bone with a severe phenotype. *Journal of Bone and Mineral Research* 2006, 21, 1136-1145.
170. Cavey, J. R.; Ralston, S. H.; Hocking, L. J.; Sheppard, P. W.; Ciani, B.; Searle, M. S.; Layfield, R., Loss of ubiquitin-binding associated with Paget's disease of bone p62 (SQSTM1) mutations. *Journal of Bone and Mineral Research* 2005, 20, 619-624.
171. Cavey, J. R.; Ralston, S. H.; Sheppard, P. W.; Ciani, B.; Gallagher, T. R. A.; Long, J. E.; Searle, M. S.; Layfield, R., Loss of ubiquitin binding is a unifying mechanism by which mutations of SQSTM1 cause Paget's disease of bone. *Calcified Tissue International* 2006, 78, 271-277.
172. Layfield, R.; Ciani, B.; Ralston, S. H.; Hocking, L. J.; Sheppard, P. W.; Searle, M. S.; Cavey, J. R., Structural and functional studies of mutations affecting the UBA domain of

SQSTM1 (p62) which cause Paget's disease of bone. *Biochemical Society Transactions* 2004, 32, 728-730.
173. Garner, T. P.; Long, J.; Layfield, R.; Searle, M. S., Impact of p62/SQSTM1 UBA Domain Mutations Linked to Paget's Disease of Bone on Ubiquitin Recognition. *Biochemistry* 2011, 50, 4665-4674.
174. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, 422, 198-207.
175. Grayson, M. A., John Bennett Fenn: A Curious Road to the Prize. *Journal of the American Society for Mass Spectrometry* 2011, 22, 1301-1308.
176. Hernandez, H.; Robinson, C. V., Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols* 2007, 2, 715-726.
177. Gabelica, V.; Vreuls, C.; Filee, P.; Duval, V.; Joris, B.; De Pauw, E., Advantages and drawbacks of nanospray for studying noncovalent protein-DNA complexes by mass spectrometry. *Rapid Communications in Mass Spectrometry* 2002, 16, 1723-1728.
178. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S., Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *Journal of the American Chemical Society* 2009, 131, 15980-+.
179. Barylyuk, K.; Balabin, R. M.; Grunstein, D.; Kikkeri, R.; Frankevich, V.; Seeberger, P. H.; Zenobi, R., What Happens to Hydrophobic Interactions during Transfer from the Solution to the Gas Phase? The Case of Electrospray-Based Soft Ionization Methods. *Journal of the American Society for Mass Spectrometry* 2011, 22, 1167-1177.
180. Yin, S.; Xie, Y. M.; Loo, J. A., Mass spectrometry of protein-ligand complexes: Enhanced gas-phase stability of ribonuclease-nucleotide complexes. *Journal of the American Society for Mass Spectrometry* 2008, 19, 1199-1208.
181. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 2006, 1, 2876-2890.
182. Venyaminov, S. Y.; Baikalov, I. A.; Shen, Z. M.; Wu, C. S. C.; Yang, J. T., Circular dichroic analysis of denatured proteins - inclusion of denatured proteins in the reference set. *Analytical Biochemistry* 1993, 214, 17-24.
183. Holzwart.G; Doty, P., Ultraviolet circular dichroism of polypeptides. *Journal of the American Chemical Society* 1965, 87, 218-&.
184. Greenfie.N; Fasman, G. D., Computed circular dichroism spectra for evaluation of protein conformation. *Biochemistry* 1969, 8, 4108-&.
185. Greenfield, N. J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* 2006, 1, 2527-2535.
186. Greenfield, N. J., Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nature Protocols* 2006, 1, 2733-2741.
187. McPhail, D.; Cooper, A., Thermodynamics and kinetics of dissociation of ligand-induced dimers of vancomycin antibiotics. *Journal of the Chemical Society-Faraday Transactions* 1997, 93, 2283-2289.
188. Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H., Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 2002, 420, 98-102.
189. Castellani, F.; van Rossum, B. J.; Diehl, A.; Rehbein, K.; Oschkinat, H., Determination of solid-state NMR structures of proteins by means of three-dimensional N-15-C-13-C-13 dipolar correlation spectroscopy and chemical shift analysis. *Biochemistry* 2003, 42, 11476-11483.
190. Clore, G. M.; Gronenborn, A. M., Determining the structures of large proteins and protein complexes by NMR. *Trends in Biotechnology* 1998, 16, 22-34.

191. Vaynberg, J.; Qin, J., Weak protein-protein interactions as probed by NMR spectroscopy. *Trends in Biotechnology* 2006, 24, 22-27.
192. Kay, L. E.; Keifer, P.; Saarinen, T., Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* 1992, 114, 10663-10665.
193. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sorensen, O. W.; Griesinger, C., A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed-field gradients. *Journal of Biomolecular Nmr* 1994, 4, 301-306.
194. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K., Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* 1997, 94, 12366-12371.
195. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by n-15 inverse detected heteronuclear NMR-spectroscopy - application to staphylococcal nuclease. *Biochemistry* 1989, 28, 8972-8979.
196. Clore, G. M.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M., Analysis of the backbone dynamics of interleukin-1-beta using 2-dimensional inverse detected heteronuclear n-15-H-1 NMR-spectroscopy. *Biochemistry* 1990, 29, 7387-7401.
197. Treweek, T. M.; Rekas, A.; Walker, M. J.; Carver, J. A., A quantitative NMR spectroscopic examination of the flexibility of the C-terminal extensions of the molecular chaperones, alpha A- and alpha B-crystallin. *Experimental Eye Research* 2010, 91, 691-699.
198. Renner, C.; Schleicher, M.; Moroder, L.; Holak, T. A., Practical aspects of the 2D N-15-{H-1}-NOE experiment. *Journal of Biomolecular Nmr* 2002, 23, 23-33.
199. Gong, Q. G.; Ishima, R., N-15-{H-1} NOE experiment at high magnetic field strengths. *Journal of Biomolecular Nmr* 2007, 37, 147-157.
200. Schanda, P.; Brutscher, B., Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society* 2005, 127, 8014-8015.
201. Schanda, P.; Van Melckebeke, H.; Brutscher, B., Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society* 2006, 128, 9042-9043.
202. Kay, L. E.; Marion, D.; Bax, A., Practical aspects of 3D heteronuclear NMR of proteins. *Journal of Magnetic Resonance* 1989, 84, 72-84.
203. Marion, D.; Kay, L. E.; Sparks, S. W.; Torchia, D. A.; Bax, A., 3-dimensional heteronuclear NMR of N-15-labeled proteins. *Journal of the American Chemical Society* 1989, 111, 1515-1517.
204. Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M., Overcoming the overlap problem in the assignment of H-1-NMR spectra of larger proteins by use of 3-dimensional heteronuclear H-1-N-15 hartmann-hahn multiple quantum coherence and nuclear overhauser multiple quantum coherence spectroscopy - application to interleukin-1-beta. *Biochemistry* 1989, 28, 6150-6156.
205. Grzesiek, S.; Bax, A., An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance* 1992, 99, 201-207.
206. Grzesiek, S.; Bax, A., Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* 1992, 114, 6291-6293.

207. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A., 3-Dimensional triple-resonance nmr-spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance* 1990, 89, 496-514.
208. Grzesiek, S.; Bax, A., Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. *Journal of Magnetic Resonance* 1992, 96, 432-440.
209. Higman, V. A. Protein NMR - A Practical Guide. http://www.protein-nmr.org.uk/spectra.html. (accessed 26-09-2012).
210. Muhandiram, D. R.; Kay, L. E., Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity. *Journal of Magnetic Resonance Series B* 1994, 103, 203-216.
211. Clubb, R. T.; Thanabal, V.; Wagner, G., A constant-time 3-dimensional triple-resonance pulse scheme to correlate intraresidue H-1(N), N-15, and C-13(') chemical-shifts in N-15-C-13-labeled proteins. *Journal of Magnetic Resonance* 1992, 97, 213-217.
212. Vranken, W.; Boucher, W.; Stevens, T.; Fogh, R.; Pajon, A.; Llinas, P.; Ulrich, E.; Markley, J.; Ionides, J.; Laue, E., The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins-structure function and bioinformatics*, 2005; Vol. 59(4), pp 687-696.
213. Orkin, S., Molecular-cloning - a laboratory manual, 2nd edition - Sambrook,J, Fritsch,EF, Maniatis,T. *Nature* 1990, 343, 604-605.
214. Sanger, F.; Nicklen, S.; Coulson, A. R., DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 1977, 74, 5463-5467.
215. Deboer, H. A.; Comstock, L. J.; Vasser, M., The tac promoter - a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 1983, 80, 21-25.
216. Tabor, S.; Richardson, C. C., A bacteriophage-t7 RNA-polymerase promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the United States of America* 1985, 82, 1074-1078.
217. Miroux, B.; Walker, J. E., Over-production of proteins in Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* 1996, 260, 289-298.
218. Dumon-Seignovert, L.; Cariot, G.; Vuillard, L., The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification* 2004, 37, 203-206.
219. Wagner, G.; Thanabal, V.; Stockman, B. J.; Peng, J. W.; Nirmala, N. R.; Hyberts, S. G.; Goldberg, M. S.; Detlefsen, D. J.; Clubb, R. T.; Adler, M., NMR-studies of structure and dynamics of isotope enriched proteins. *Biopolymers* 1992, 32, 381-390.
220. Gill, S. C.; Vonhippel, P. H., Calculation of protein extinction coefficients from amino-acid sequence data. *Analytical Biochemistry* 1989, 182, 319-326.
221. Kaltashov, I. A.; Mohimen, A., Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Analytical Chemistry* 2005, 77, 5370-5379.
222. Veros, C. T.; Oldham, N. J., Quantitative determination of lysozyme-ligand binding in the solution and gas phases by electrospray ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry* 2007, 21, 3505-3510.
223. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *Journal of Biomolecular Nmr* 1992, 2, 661-665.
224. Hwang, T. L.; Shaka, A. J., Water suppression that works - excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance Series A* 1995, 112, 275-279.
225. Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K., Recommendations for the presentation of NMR structures of

proteins and nucleic acids - IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *European Journal of Biochemistry* 1998, 256, 1-15.
226. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R., Investigation of exchange processes by 2-dimensional NMR-spectroscopy. *Journal of Chemical Physics* 1979, 71, 4546-4553.
227. Braunschweiler, L.; Ernst, R. R., Coherence transfer by isotropic mixing - application to proton correlation spectroscopy. *Journal of Magnetic Resonance* 1983, 53, 521-528.
228. Davis, A. L.; Keeler, J.; Laue, E. D.; Moskau, D., Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. *Journal of Magnetic Resonance* 1992, 98, 207-216.
229. Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M., Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *Journal of Magnetic Resonance* 1991, 93, 151-170.
230. Bax, A.; Davis, D. G., MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance* 1985, 65, 355-360.
231. Schleucher, J.; Sattler, M.; Griesinger, C., Coherence selection by gradients without signal attenuation - application to the 3-dimensional HNCO experiment. *Angewandte Chemie-International Edition in English* 1993, 32, 1489-1491.
232. Wu, C. J.; Conze, D. B.; Li, T.; Srinivasula, S. M.; Ashwell, J. D., NEMO is a sensor of Lys 63-linked polyubiquitination and functions in NF-kappa B activation. *Nature Cell Biology* 2006, 8, 398-U58.
233. Hadian, K.; Griesbach, R. A.; Dornauer, S.; Wanger, T. M.; Nagel, D.; Metlitzky, M.; Beisker, W.; Schmidt-Supprian, M.; Krappmann, D., NF-kappa B Essential Modulator (NEMO) Interaction with Linear and Lys-63 Ubiquitin Chains Contributes to NF-kappa B Activation. *Journal of Biological Chemistry* 2011, 286, 26107-26117.
234. Kozlov, G.; Peschard, P.; Zimmerman, B.; Lin, T.; Moldoveanu, T.; Mansur-Azzam, N.; Gehring, K.; Park, M., Structural basis for UBA-mediated dimerization of c-Cbl ubiquitin ligase. *Journal of Biological Chemistry* 2007, 282, 27547-27555.
235. Peschard, P.; Kozlov, G.; Lin, T.; Mirza, A.; Berghuis, A. M.; Lipkowitz, S.; Park, M.; Gehring, K., Structural basis for ubiquitin-mediated dimerization and activation of the ubiquitin protein ligase Cbl-b. *Molecular Cell* 2007, 27, 474-485.
236. Tse, M. K.; Hui, S. K.; Yang, Y. H.; Yin, S. T.; Hu, H. Y.; Zou, B.; Wong, B. C. Y.; Sze, K. H., Structural Analysis of the UBA Domain of X-linked Inhibitor of Apoptosis Protein Reveals Different Surfaces for Ubiquitin-Binding and Self-Association. *Plos One* 2011, 6.
237. Bayrer, J. R.; Zhang, W.; Weiss, M. A., Dimerization of doublesex is mediated by a cryptic ubiquitin-associated domain fold - Implications for sex-specific gene regulation. *Journal of Biological Chemistry* 2005, 280, 32989-32996.
238. Gyrd-Hansen, M.; Darding, M.; Miasari, M.; Santoro, M. M.; Zender, L.; Xue, W.; Tenev, T.; da Fonseca, P. C. A.; Zvelebil, M.; Bujnicki, J. M.; Lowe, S.; Silke, J.; Meier, P., IAPs contain an evolutionarily conserved ubiquitin-binding domain that regulates NF-kappa B as well as cell survival and oncogenesis. *Nature Cell Biology* 2008, 10, 1309-U130.
239. Dey, S.; Pal, A.; Chakrabarti, P.; Janin, J., The Subunit Interfaces of Weakly Associated Homodimeric Proteins. *Journal of Molecular Biology* 2010, 398, 146-160.
240. Pelton, J. T.; McLean, L. R., Spectroscopic methods for analysis of protein secondary structure. *Analytical Biochemistry* 2000, 277, 167-176.
241. Garner, T. Structural and biophysical investigations into ubiquitin binding proteins. University of Nottingham, University of Nottingham, 2010.

242. Konermann, L.; Douglas, D. J., Unfolding of proteins monitored by electrospray ionization mass spectrometry: A comparison of positive and negative ion modes. *Journal of the American Society for Mass Spectrometry* 1998, 9, 1248-1254.
243. Hopper, J. T. S.; Sokratous, K.; Oldham, N. J., Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure. *Analytical Biochemistry* 2012, 421, 788-790.
244. Grasberger, B. L.; Gronenborn, A. M.; Clore, G. M., The analysis of the backbone dynamics of interleukin-8 by N-15 relaxation measurements. *Journal of Molecular Biology* 1993, 230, 364-372.
245. Zhou, Z. R.; Gao, H. C.; Zhou, C. J.; Chang, Y. G.; Hong, J.; Song, A. X.; Lin, D. H.; Hu, H. Y., Differential ubiquitin binding of the UBA domains from human c-Cbl and Cbl-b: NMR structural and biochemical insights. *Protein Science* 2008, 17, 1805-1814.
246. Cohen, P., The origins of protein phosphorylation. *Nature Cell Biology* 2002, 4, E127-E130.
247. Burnett, G.; Kennedy, E. P., The enzymatic phosphorylation of proteins. *Journal of Biological Chemistry* 1954, 211.
248. Pearlman, S. M.; Serber, Z.; Ferrell, J. E., Jr., A Mechanism for the Evolution of Phosphorylation Sites. *Cell* 2011, 147.
249. Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T., Protein tyrosine phosphatases in the human genome. *Cell* 2004, 117.
250. Holt, L. J.; Tuch, B. B.; Villen, J.; Johnson, A. D.; Gygi, S. P.; Morgan, D. O., Global Analysis of Cdk1 Substrate Phosphorylation Sites Provides Insights into Evolution. *Science* 2009, 325.
251. Nash, P.; Tang, X. J.; Orlicky, S.; Chen, Q. H.; Gertler, F. B.; Mendenhall, M. D.; Sicheri, F.; Pawson, T.; Tyers, M., Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 2001, 414, 514-521.
252. Wagner, L. E.; Li, W. H.; Joseph, S. K.; Yule, D. I., Functional consequences of phosphomimetic mutations at key cAMP-dependent protein kinase phosphorylation sites in the type 1 inositol 1,4,5-trisphosphate receptor. *Journal of Biological Chemistry* 2004, 279, 46242-46252.
253. Pondugula, S. R.; Brimer-Cline, C.; Wu, J.; Schuetz, E. G.; Tyagi, R. K.; Chen, T. S., A Phosphomimetic Mutation at Threonine-57 Abolishes Transactivation Activity and Alters Nuclear Localization Pattern of Human Pregnane X Receptor. *Drug Metabolism and Disposition* 2009, 37, 719-730.
254. Makino, Y.; Johnson, R. C.; Yu, Y. L.; Takamiya, K.; Huganir, R. L., Enhanced synaptic plasticity in mice with phosphomimetic mutation of the GluA1 AMPA receptor. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108, 8450-8455.
255. Potter, M. D.; Barbero, S.; Cheresh, D. A., Tyrosine phosphorylation of VE-cadherin prevents binding of p120- and beta-catenin and maintains the cellular mesenchymal state. *Journal of Biological Chemistry* 2005, 280, 31906-31912.
256. Pecina, P.; Borisenko, G. G.; Belikova, N. A.; Tyurina, Y. Y.; Pecinova, A.; Lee, I.; Samhan-Arias, A. K.; Przyklenk, K.; Kagan, V. E.; Huttemann, M., Phosphomimetic Substitution of Cytochrome c Tyrosine 48 Decreases Respiration and Binding to Cardiolipin and Abolishes Ability to Trigger Downstream Caspase Activation. *Biochemistry* 2010, 49, 6705-6714.
257. Matsumoto, G.; Wada, K.; Okuno, M.; Kurosawa, M.; Nukina, N., Serine 403 Phosphorylation of p62/SQSTM1 Regulates Selective Autophagic Clearance of Ubiquitinated Proteins. *Molecular cell* 2011, 44, 279-89.

258. Ohguro, H.; Vanhooser, J. P.; Milam, A. H.; Palczewski, K., Rhodopsin phosphorylation and dephosphorylation *in vivo*. *Journal of Biological Chemistry* 1995, 270, 14259-14262.
259. Wild, P.; Farhan, H.; McEwan, D. G.; Wagner, S.; Rogov, V. V.; Brady, N. R.; Richter, B.; Korac, J.; Waidmann, O.; Choudhary, C.; Dotsch, V.; Bumann, D.; Dikic, I., Phosphorylation of the Autophagy Receptor Optineurin Restricts Salmonella Growth. *Science* 2011, 333, 228-233.
260. Linares, J. F.; Amanchy, R.; Diaz-Meco, M. T.; Moscat, J., Phosphorylation of p62 by cdk1 Controls the Timely Transit of Cells through Mitosis and Tumor Cell Proliferation. *Molecular and Cellular Biology* 2011, 31, 105-117.
261. Rajesh, S.; Bago, R.; Odintsova, E.; Muratov, G.; Baldwin, G.; Sridhar, P.; Rajesh, S.; Overduin, M.; Berditchevski, F., Binding to Syntenin-1 Protein Defines a New Mode of Ubiquitin-based Interactions Regulated by Phosphorylation. *Journal of Biological Chemistry* 2011, 286, 39606-39614.
262. Behrends, C.; Sowa, M. E.; Gygi, S. P.; Harper, J. W., Network organization of the human autophagy system. *Nature* 2010, 466, 68-U84.
263. Mizushima, N., Autophagy: process and function. *Genes & Development* 2007, 21, 2861-2873.
264. Levine, B.; Kroemer, G., Autophagy in the pathogenesis of disease. *Cell* 2008, 132, 27-42.
265. Levine, B.; Mizushima, N.; Virgin, H. W., Autophagy in immunity and inflammation. *Nature* 2011, 469, 323-335.
266. Noda, T.; Ohsumi, Y., Tor, a phosphatidylinositol kinase homologue, controls autophagy in yeast. *Journal of Biological Chemistry* 1998, 273, 3963-3966.
267. Hamasaki, M.; Yoshimori, T., Where do they come from? Insights into autophagosome formation. *Febs Letters* 2010, 584, 1296-1301.
268. Kroemer, G.; Jaattela, M., Lysosomes and autophagy in cell death control. *Nature Reviews Cancer* 2005, 5, 886-897.
269. Weidberg, H.; Shpilka, T.; Shvets, E.; Abada, A.; Shimron, F.; Elazar, Z., LC3 and GATE-16 N Termini Mediate Membrane Fusion Processes Required for Autophagosome Biogenesis. *Developmental Cell* 2011, 20, 444-454.
270. Kouno, T.; Mizuguchi, M.; Tanida, I.; Ueno, T.; Kanematsu, T.; Mori, Y.; Shinoda, H.; Hirata, M.; Kominami, E.; Kawano, K., Solution structure of microtubule-associated protein light chain 3 and identification of its functional subdomains. *Journal of Biological Chemistry* 2005, 280, 24610-24617.
271. Satoo, K.; Noda, N. N.; Kumeta, H.; Fujioka, Y.; Mizushima, N.; Ohsumi, Y.; Inagaki, F., The structure of Atg4B-LC3 complex reveals the mechanism of LC3 processing and delipidation during autophagy. *Embo Journal* 2009, 28, 1341-1350.
272. Pandey, U. B.; Nie, Z. P.; Batlevi, Y.; McCray, B. A.; Ritson, G. P.; Nedelsky, N. B.; Schwartz, S. L.; DiProspero, N. A.; Knight, M. A.; Schuldiner, O.; Padmanabhan, R.; Hild, M.; Berry, D. L.; Garza, D.; Hubbert, C. C.; Yao, T. P.; Baehrecke, E. H.; Taylor, J. P., HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. *Nature* 2007, 447, 859-863.
273. Korolchuk, V. I.; Mansilla, A.; Menzies, F. M.; Rubinsztein, D. C., Autophagy Inhibition Compromises Degradation of Ubiquitin-Proteasome Pathway Substrates. *Molecular Cell* 2009, 33, 517-527.
274. Bjorkoy, G.; Lamark, T.; Pankiv, S.; Overvatn, A.; Brech, A.; Johansen, T., Monitoring autophagic degradation of p62/SQSTM1. In *Methods in Enzymology: Autophagy in Mammalian Systems, Vol 452, Pt B*, Elsevier Academic Press Inc: San Diego, 2009; Vol. 452, pp 181-197.

275. Noda, N. N.; Kumeta, H.; Nakatogawa, H.; Satoo, K.; Adachi, W.; Ishii, J.; Fujioka, Y.; Ohsumi, Y.; Inagaki, F., Structural basis of target recognition by Atg8/LC3 during selective autophagy. *Genes to Cells* 2008, 13, 1211-1218.
276. Kirkin, V.; McEwan, D. G.; Novak, I.; Dikic, I., A Role for Ubiquitin in Selective Autophagy. *Molecular Cell* 2009, 34, 259-269.
277. Johansen, T.; Lamark, T., Selective autophagy mediated by autophagic adapter proteins. *Autophagy* 2011, 7, 279-296.
278. Itakura, E.; Mizushima, N., p62 targeting to the autophagosome formation site requires self-oligomerization but not LC3 binding. *Journal of Cell Biology* 2011, 192, 17-27.
279. Kirkin, V.; Lamark, T.; Johansen, T.; Dikic, I., NBR1 cooperates with p62 in selective autophagy of ubiquitinated targets. *Autophagy* 2009, 5, 732-733.
280. Rozenknop, A.; Rogov, V. V.; Rogova, N. Y.; Lohr, F.; Guntert, P.; Dikic, I.; Dotsch, V., Characterization of the Interaction of GABARAPL-1 with the LIR Motif of NBR1. *Journal of Molecular Biology* 2011, 410, 477-487.
281. Sollenberger, J. Biophysical studies of the molecular interactions of autophagy marker $LC3_B$. University of Nottingham, 2011.
282. Garner, T. *Structural study of the human scaffold protein p62: insight into the binding of ubiquitin by the UBA domain and N-terminal regions*; University of Nottingham: 2007.
283. Chang, S. W.; Kim, J. H.; Shin, J., p62 forms a ternary complex with PKC zeta and PAR-4 and antagonizes PAR-4-induced PKC zeta inhibition. *Febs Letters* 2002, 510, 57-61.
284. Thurston, T. L. M.; Ryzhakov, G.; Bloor, S.; von Muhlinen, N.; Randow, F., The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria. *Nature Immunology* 2009, 10, 1215-U103.
285. Felli, I. C.; Pierattelli, R., Recent progress in NMR spectroscopy: Toward the study of intrinsically disordered proteins of increasing size and complexity. *Iubmb Life* 2012, 64, 473-481.
286. Novacek, J.; Haba, N. Y.; Chill, J. H.; Zidek, L.; Sklenar, V., 4D Non-uniformly sampled HCBCACON and (1) J(NC alpha)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *Journal of Biomolecular Nmr* 2012, 53, 139-148.
287. Sims, J. J.; Cohen, R. E., Linkage-Specific Avidity Defines the Lysine 63-Linked Polyubiquitin-Binding Preference of Rap80. *Molecular Cell* 2009, 33, 775-783.
288. Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry* 1996, 68, 1-8.

Appendix A : Primer sequences

T414K

5′ GGC GGC TGG CTC AAA AGG CTC CTG CAG ACC 3′

3′ GGT CTG CAG GAG CCT TTT GAG CCA GCC GCC 5′

R415K

5′ GGC GGC TGG CTC ACC AAA CTC CTG CAG ACC 3′

3′ GGT CTG CAG GAG AAA GGT GAG CCA GCC GCC 5′

Q418A

5′ GG CTC ACC AGG CTC CTG AAA ACC AAG AAC TAT GAC ATC GG 3′

3′ CC GAT GTC ATA GTT CTT GGT TTT CAG GAG CCT GGT GAG CC 5′

T419K

5′ C ACC AGG CTC CTG CAG AAA AAG AAC TAT GAC ATC G 3′

3′ C GAT GTC ATA GTT CTT TTT CTG CAG GAG CCT GGT G 5′

S403A

5′ CC CTC TCC CAG ATG CTG GCG ATG GGC TTC TCT GAT GAA GGC 3′

3′ GCC TTC ATC AGA GAA GCC CAT CGC CAG CAT CTG GGA GAG GG 5′

S403D

5′ CC CTC TCC CAG ATG CTG GAT ATG GGC TTC TCT GAT GAA GGC 3′

3′ GCC TTC ATC AGA GAA GCC CAT ATC CAG CAT CTG GGA GAG GG 5′

S403E

5′ CC CTC TCC CAG ATG CTG GAG ATG GGC TTC TCT GAT GAA G 3′

3′ C TTC ATC AGA GAA GCC CAT CTC CAG CAT CTG GGA GAG GG 5′